



TESIS DOCTORAL

Modelo teórico-metodológico basado en el KDD para la integración y explotación de datos bibliográficos de patentes

Autora:

María de Jesús Madera Jaramillo

Directores:

Elías Sanz Casado

Salvador Gorbea Portal

DEPARTAMENTO DE BIBLIOTECONOMÍA Y DOCUMENTACIÓN

Getafe, junio de 2014



Universidad
Carlos III de Madrid
www.uc3m.es



TESIS DOCTORAL

Modelo teórico-metodológico basado en el KDD para la integración y explotación de datos bibliográficos de patentes

Autora: *María de Jesús Madera Jaramillo*

Directores: Elías Sanz Casado

Salvador Gorbea Portal

Firma del Tribunal Calificador:

Firma

Presidente: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Secretario: (Nombre y apellidos)

Calificación:

Getafe, de de



Universidad
Carlos III de Madrid
www.uc3m.es

El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad.

Víctor Hugo (1802-1885)

Agradecimientos

A la Universidad Nacional Autónoma de México, y a la Fundación Carolina por el apoyo financiero para la realización de este proyecto.

Al Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid, por el apoyo en infraestructura durante mis estancias en la universidad.

A la Dra. María Luisa Lascurain, al Dr. Salvador Gorbea Portal, y al Dr. Elías Sanz Casado por sus acertadas recomendaciones durante la elaboración de la tesis.

A la Dra. Carmen Martín Moreno por su ayuda con los trámites administrativos, y por el apoyo moral.

A mis amigos y familia, por hacerme más divertido este trayecto de mi vida.

Tabla de contenido

1. Introducción	1
1.1. Descubrimiento de conocimiento en bases de datos	3
1.2. <i>Data Warehouse</i>	5
1.3. Minería de datos	7
1.4. La evaluación de la ciencia	10
1.4.1. Opinión de expertos	12
1.4.2. Indicadores bibliométricos	12
1.4.2.1. De productividad científica	12
1.4.2.2. Indicadores de conexión o nexos informativos entre trabajos y autores científicos	14
1.4.2.3. Indicadores del impacto de los trabajos	14
1.4.2.4. Indicador de impacto de las revistas	14
1.4.2.5. Indicadores de visibilidad de las revistas	15
1.5. El estudio bibliométrico de las patentes	15
1.5.1. Patentes	16
1.5.2. Fundamentos jurídicos de las patentes	16
1.5.2.1. Proceso de protección de las patentes	18
1.5.3. Fundamentos económicos de las patentes	22
1.5.4. Documentos de patentes	23
1.5.5. Temas de Investigación sobre las patentes	27
1.5.5.1. Rendimiento tecnológico	27
1.5.5.2. Tecnologías emergentes	27
1.5.5.3. Difusión del conocimiento y dinámicas del cambio técnico	27
1.5.5.4. Geografía de la invención	28
1.5.5.5. El valor económico de las invenciones	28
1.5.5.6. Rendimiento y movilidad de los investigadores	28
1.5.5.7. El papel de las universidades en el desarrollo tecnológico	28
1.5.5.8. Globalización de las actividades de I+D	29
1.5.5.9. Estrategias de las empresas en materia de patentes	29

Tabla de contenido

1.5.6. Indicadores de patentes	29
1.5.6.1. Indicadores relacionados con las fechas de referencia	30
1.5.6.2. Indicadores relacionados con el país de referencia	30
1.5.6.3. Indicadores sobre Familia de patentes	31
1.5.6.4. Indicadores sobre la clasificación de las patentes	33
1.5.6.5. Indicadores basados de referencias de patentes	34
1.5.6.6. Indicadores del valor de una patente	34
2. Hipótesis y objetivos	37
2.1. Problema de investigación	39
2.2. Hipótesis	39
2.3. Objetivos	40
2.3.1. Objetivo general	40
2.3.2. Objetivos Específicos	40
3. Determinación de fuentes de datos y herramientas para la obtención y procesamiento de indicadores de patentes	43
3.1. Modelo de integración de datos	45
3.1.1. Fuente de datos	45
3.1.1.1. <i>Interpat</i> de INVENES	45
3.1.1.2. <i>Derwent Innovation Index</i>	49
3.1.1.3. El Servicio europeo <i>esp@cenet</i>	56
3.1.2. Modelo de <i>Data Warehouse</i>	58
3.2. Tratamiento de los datos	60
3.3. Visualización de datos	62
3.3.1. <i>Gephi</i>	62
3.3.2. <i>Pajek</i>	64
3.3.3. <i>VOSviewer</i>	65
4. Diseño e implementación de un modelo teórico-metodológico para la integración de datos bibliográficos de patentes y la obtención de sus indicadores	67
4.1. Modelo teórico-metodológico de integración de datos bibliográficos de patentes	69
4.1.1. Selección de datos	69
4.1.2. Modelo de Integración de bases de datos	70

4.1.3 Definición de relaciones y atributos	75
4.2. Tratamiento de datos	91
4.2.1. Carga de los datos	91
4.2.1.1. Catálogos de autoridad	96
4.2.1.2. Reetiquetado de datos	98
4.2.1.3. Ingresar datos <i>Interpat</i>	101
4.2.1.4. Generación de lista con números de patentes <i>Interpat</i>	102
4.2.1.5. Ingresar datos <i>Derwent Innovation Index</i>	105
4.2.1.6. Generación de lista con números de patentes citadas	108
4.2.1.7. Ingresar datos <i>Worldwide-espacenet</i>	110
4.2.1.8. Procesa patentes referenciadas	111
4.2.2. Normalización	113
4.2.2.1. Normalizar nombre del solicitantes	113
4.2.2.2. Codificación del solicitantes	114
4.2.3. Generación de reporte	116
4.2.3.1. Evolución general de patentes	119
4.2.3.2 Distribución general y evolución por sectores	121
4.2.3.3. Distribución general y evolución por temática a nivel clase	122
4.2.3.4. Distribución general y evolución por temática nivel subgrupo	124
4.2.3.5. Distribución por sector y temática a nivel clase y subgrupo	125
4.2.3.6. Evolución general de las familias de patentes	128
4.2.3.7. Evolución general de las familias de patente por Tamaño	129
4.2.3.8. Temática de las familias de patentes a nivel clase y subgrupo	130
4.2.3.9. Distribución de la nacional de las patentes referenciadas	133
4.2.4.0. Temática de las patentes referenciadas a nivel clase y subgrupo	135
4.2.4.1. Distribución de las patentes referenciadas por temática a nivel clase y nacionalidad	136

Tabla de contenido

4.2.4.2. Distribución de la temática de las patentes fuente y temática de las patentes referenciadas	138
4.2.4.3. Evolución de las familias de patentes triádicas	139
4.2.4.4. Distribución de las familias de patentes triádicas por área temática	140
4.3. Aplicación del Modelo teórico-metodológico: Estudio de caso con una muestra de patentes españolas	142
4.3.1 Capacidad innovadora de la Comunidad de Madrid a partir de las patentes concedidas entre 1996 y 2007	142
4.3.1.1 Las invenciones en la CM y en el conjunto de España	143
4.3.1.2 Características de la invención en la CM	144
4.3.2. Origen del conocimiento tecnológico de las patentes españolas	153
4.3.3. Interés de las invenciones españolas a partir del estudio de las familias de patentes	165
5. Conclusiones	173
6. Referencias	179
7. Anexo	199

Lista de tablas

Tabla 3.1: Lista de etiquetas correspondientes a la base de datos <i>Interpat</i>	48
Tabla 3.2: Lista de etiquetas correspondientes a la base de datos <i>Derwent</i>	51
Tabla 3.3: Lista de etiquetas presentes en los registros a la base de datos <i>Worldwide-espacenet</i>	58
Tabla 3.4: Gephi: Categorías y Algoritmos	63
Tabla 4.1: Definición de relaciones y atributos	75
Tabla 4.2: Fragmento de un archivo de tipo texto sin formato, fecha de descarga 2008	93
Tabla 4.3: Fragmento de un archivo de tipo texto sin formato, fecha de descarga 2010	94
Tabla 4.4: Fragmento de un archivo de tipo texto reetiquetado	100
Tabla 4.5: Fragmento de un archivo de la base de datos <i>Derwent</i> <i>Innovation Index</i>	105
Tabla 4.6: Número de patentes concedidas en España y en la CM	143
Tabla 4.7: Evolución de patentes en la CM según sectores institucionales e incrementos anuales (entre paréntesis)	144
Tabla 4.8: Temática de las patentes concedidas a la CM	148
Tabla 4.9: Temática de las patentes concedidas a la CM a nivel clase según código CIP	152
Tabla 4.10: Distribución de patentes fuente vs. Referenciadas, núm. índice	158
Tabla 4.11: Origen geográfico de las referencias incluidas en las patentes	159
Tabla 4.12: Distribución de la temática de las patentes fuente a nivel sección	160
Tabla 4.13: Distribución de la temática de las patentes referenciadas a nivel sección	161
Tabla 4.14: Temática patentes fuente vs. referenciadas	161
Tabla 4.15: Temática patentes fuente vs. referenciadas a nivel clase	162
Tabla 4.16: Temática patentes fuente vs. país de patente referenciada	164
Tabla 4.17: Evolución temporal de las patentes prioritarias	166
Tabla 4.18: Tamaño de las familias de patentes	167
Tabla 4.19: Distribución de la temática de las familias de patentes a nivel sección	168
Tabla 4.20: Distribución de la temática de las familias de patentes a nivel sección vs País de protección	169

Tabla 4.21: Evolución temporal de las familias de patentes triádicas	171
Tabla 4.22: Temática a primer nivel (CIP) de las familias de patentes triádicas	172

Lista de figuras

Figura 1.1: Proceso de descubrimiento de conocimiento en bases de datos	5
Figura 1.2: Evolución de las tecnologías de Bases de Datos	8
Figura 1.3: Arquitectura de un sistema de minería de datos	9
Figura 1.4: Relaciones interdisciplinarias de la Minería de Datos	10
Figura 1.5: Proceso de solicitud de patente mediante el PCT.	22
Figura 1.6: Primera página de una patente. Fuente: Base de datos INVENES	26
Figura 3.1: Pantalla de búsqueda avanzada en la interfaz <i>Interpat</i>	47
Figura 3.2: Pantalla de búsqueda experta en la interfaz <i>Interpat</i>	47
Figura 3.3: Pantalla de búsqueda en la interfaz de <i>Derwent Innovation Index</i>	50
Figura 3.4: Pantalla de búsqueda de la interfaz <i>espacenet</i>	57
Figura 3.5: Funcionamiento de los <i>scripts</i> CGI	61
Figura 3.6: Herramienta de visualización <i>Gephi</i>	63
Figura 3.7: Herramienta de visualización <i>Pajek</i>	65
Figura 3.8: Herramienta <i>VOSviewer</i>	66
Figura 4. 1: Modelo de integración de datos bibliográficos de patentes	72
Figura 4.2: Diseño de la base de datos	74
Figura 4.3: Diagrama general de tratamiento de datos	95
Figura 4.4: Página inicial	96
Figura 4.5: Catálogo de autoridad de atributos <i>Interpat</i>	97
Figura 4.6: Catálogo de autoridad de la Clasificación Internacional de Patentes	98
Figura 4.7: Diagrama de secuencia de reetiquetado	100
Figura 4.8: Diagrama de secuencia para procesar datos <i>Interpat</i>	102
Figura 4.9: Diagrama de secuencia para generar lista de números de patentes	103
Figura 4.10: Lista de líneas de búsqueda de números de patentes	104
Figura 4.11: Interfaz de búsqueda de <i>Derwent Innovation Index</i>	104
Figura 4.12: Diagrama de secuencia para procesar archivos <i>Derwent Innovation Index</i>	107
Figura 4.13: Diagrama de secuencia para generar lista de patentes referenciadas	109
Figura 4.14: Lista de líneas de búsqueda de patentes referenciadas	109
Figura 4.15: Fragmento de patentes citadas, formato <i>csv</i>	110
Figura 4.16: Diagrama de secuencia para ingresar patentes referenciadas	111

Lista de figuras

Figura 4.17: Diagrama de secuencia para ingresar patentes referenciadas	112
Figura 4.18: Diagrama de secuencia para normalizar nombre de solicitantes	114
Figura 4.19: Diagrama de secuencia para codificar sector de solicitantes	116
Figura 4.20: Diagrama de secuencia general	118
Figura 4.21: Página de parámetros	119
Figura 4.22: Diagrama de secuencia para generar evolución general de patentes	120
Figura 4.23: Pantalla con reporte de evolución general de patentes	120
Figura 4.24: Diagrama de secuencia para generar evolución general por sector	122
Figura 4.25: Diagrama de secuencia para generar evolución general por temática a nivel clase	124
Figura 4.26: Diagrama de secuencia para generar evolución general por temática a nivel subgrupo	125
Figura 4.27: Diagrama de secuencia para generar distribución por sector y temática nivel clase	127
Figura 4.28: Diagrama de secuencia para generar distribución por sector y temática a nivel subgrupo	127
Figura 4.29: Diagrama de secuencia para generar evolución de las familias de patentes	129
Figura 4.30: Diagrama de secuencia para generar evolución de las familias de patentes por tamaño	130
Figura 4.31: Diagrama de secuencia para generar evolución de las familias de patentes por área temática a nivel clase	132
Figura 4.32: Diagrama de secuencia para generar evolución de las familias de patentes por área temática a nivel clase	133
Figura 4.33: Diagrama de secuencia de la distribución por nacionalidad de las patentes referenciadas	134
Figura 4.34: Diagrama de secuencia para generar evolución de la temática referenciada a nivel clase	135
Figura 4.35: Diagrama de secuencia para generar evolución de la temática referenciada a nivel subgrupo	136
Figura 4.36: Diagrama de secuencia para generar distribución de las patentes referenciadas por temática y nacionalidad	137
Figura 4.37: Diagrama de secuencia temática fuente y temática referenciada	139

Figura 4.38: Diagrama de secuencia de evolución de familias triádicas	140
Figura 4.39: Diagrama de secuencia de familias triádicas por área temática	141
Figura 4.40: Incremento del número de patentes concedidas por sectores (números índice. Año base 1996)	146
Figura 4.41: Evolución de la actividad patentadora de los sectores (análisis de correspondencias)	147
Figura 4.42: Evolución temporal de los temas en los que se patenta en la CM (análisis de correspondencias)	149
Figura 4.43: Análisis de correspondencias. Temas (a nivel sección) vs sectores	150
Figura 4.44: Análisis de correspondencias. Temas (nivel 3) vs sectores	151
Figura 4.45: Formulario de selección de parámetros para obtener la evolución de las patentes fuente	155
Figura 4.46: Lista de reportes de patentes Derwent	155
Figura 4.47: Reporte de la evolución temporal de las patentes fuente	156
Figura 4.48: Evolución de las patentes fuente (1997-2008)	157
Figura 4.49: Evolución general de las patentes referenciadas	158
Figura 4.50: Distribución anual de las patentes fuente y de las referenciadas. Números índice	159
Figura 4.51: Temática de las patentes fuentes y de las patentes referenciadas. Nivel sección de la CIP	162
Figura 4.52: Temas de las patentes fuentes y de las patentes referenciadas. Nivel clase de la CIP	163
Figura 4.53: Temas vs. Países de las patentes referenciadas. AC	164
Figura 4.54: Evolución temporal patentes prioritarias	166
Figura 4.55: Temática a primer nivel (CIP) de las familias de patentes	168
Figura 4.56: Temática de las familias de patentes vs País	169
Figura 4.57: Temática de las familias de patentes vs países de las familias de patentes (20 o más miembros)	170
Figura 4.58: Evolución temporal de las familias de patentes triádicas	171

CAPÍTULO 1

INTRODUCCIÓN

1. Introducción

En la actualidad la mayoría de las empresas e instituciones académicas hacen uso de las tecnologías de la información para difundir en forma rápida, fácil y a bajo costo sus avances tecnológicos. Esto se debe a que las tecnologías de la Información son herramientas que proporcionan los mecanismos que apoyan el proceso de difusión, y además son un apoyo para que las personas tomen decisiones en forma eficiente y efectiva. Algunas de estas herramientas son *Internet*, *Data Warehouse* (repositorios de información), la minería de datos, de texto, el proceso de descubrimiento de conocimiento en bases de datos, entre otras.

Una de las áreas de estudio que no se ha quedado atrás en el uso de estas tecnologías es la denominada como el área de Estudios Métricos de la Información, y es en este punto donde este tipo de estudios los Estudios Métricos de la Información se vincula con las Ciencias de la Computación, en particular en el campo de Bases de Datos, *Data Warehouse*, Minería de Datos y el proceso conocido como Descubrimiento de Conocimiento en Bases de Datos o KDD (por sus siglas en inglés *Knowledge Discovery in Databases*).

1.1. Descubrimiento de conocimiento en bases de datos

El Descubrimiento de Conocimiento en Bases de Datos (KDD) es un conjunto de procesos introducidos al análisis de las Bases de datos a finales de los años 80 (Swanson 1986, 1989, citado por Hudomali y Vidmar (2003)). Estos autores señalan que este proceso ha sido definido por Fayyad et al. (1996) como *procesos no comunes de identificación de patrones válidos, novedoso, potencialmente útiles y finalmente comprensibles en los datos*. Su propia definición como proceso se debe a que esta técnica entraña un conjunto de pasos los cuales han sido definidos por Fayyad, Piatetsky-Shapiro y Smyth (1996), así como por Han y Kamber (2001) y que comprende los procesos siguientes: limpieza de datos, integración de datos, selección y transformación de datos (*data Warehouse*), minería de datos, evaluación de patrones y presentación del conocimiento descubierto. También, Frawley et al. (1992) define al

procesos KDD como la tarea no trivial de extraer información implícita, previamente desconocida y potencialmente útil de bases de datos.

No obstante que hay varias definiciones del proceso KDD, se puede observar que todos los autores coinciden en que éste es un proceso con el cual es posible obtener información útil de grandes almacenes de datos. En la Figura 1.1 se observan los pasos involucrados en el proceso de descubrir conocimiento en bases de datos los cuales se definen a continuación:

- Limpieza e integración de los datos. En esta primera etapa del proceso se realiza una selección de las fuentes de datos, éstas pueden ser bases de datos y/o archivos. Además, se eliminan los datos inconsistentes y se combinan las diferentes fuentes de datos que fueron seleccionadas en un *Data Warehouse*.
- Selección y transformación de los datos. En esta paso del proceso se seleccionan los atributos que serán utilizados para el análisis y son transformados en un formato apropiado para el análisis que será realizado posteriormente con la minería de datos. Las dos primeras etapas del proceso KDD son las etapas en las que se consume más tiempo dado que es aquí donde se debe tener especial cuidado en la “limpieza” que haya en los datos, ya que sin calidad en ellos no habrá calidad en los resultados obtenidos a través de la minería de datos.
- Minería de datos. La minería de datos es la parte medular del proceso KDD y su objetivo, como se mencionó anteriormente, es identificar y extraer patrones de comportamiento descriptivo y predictivos de grandes almacenes de datos.
- Evaluación de patrones y presentación del nuevo conocimiento. Es en esta etapa del proceso donde se aplican distintas medidas, principalmente estadísticas para identificar los patrones más interesantes. También, se utilizan técnicas para visualizar los patrones descubiertos y así facilitar la interacción del usuario con el sistema.

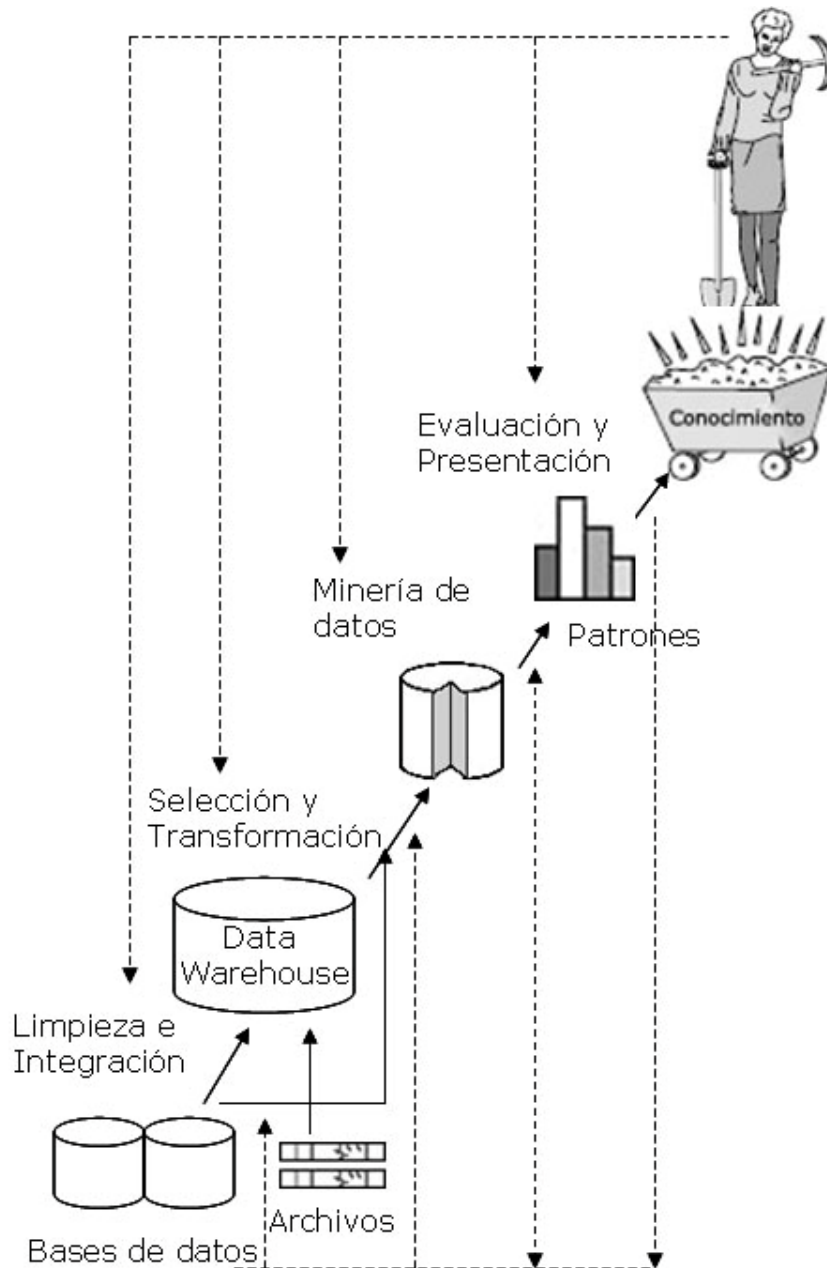


Figura 1.1: Proceso de descubrimiento de conocimiento en bases de datos.

Fuente: Han y Kamber, 2001.

1.2. *Data Warehouse*

Los *Data Warehouse* son repositorios de datos recabados de diversas fuentes, almacenados bajo un sistema unificado, y que usualmente residen en un solo sitio. Los datos se ordenan alrededor de temas principales y se usa como soporte para el proceso

de toma de decisiones gerenciales, es decir, solamente están aquellos datos que tienen el suficiente valor como para soportar una decisión, información histórica estructurada. Estos repositorios pueden ser una base de datos relacional o bien un cubo de datos multidimensional.

Existen diferentes definiciones de los *Data Warehouse*, las dos definiciones mas importantes y completas fueron propuestas por Bill Inmon y Ralph Kimball, cofundadores de data warehousing. Bill Inmon (2008) define un *Data Warehouse* como una colección de datos con carácter temático, integrada, no volátil, histórica que ayudan en la toma de decisiones. Ralph Kimball (1998) proporciona una definición más simple. Kimball dijo que un *Data Warehouse* es una copia de la data transaccional (también llamada información de las transacciones) específicamente estructurada para realizar consultas y análisis.

La data transaccional son los datos que describen un evento, es decir, el cambio como resultado de una transacción. La información transaccional siempre tiene una dimensión de tiempo, y un valor numérico, y se refiere a uno o más objetos llamados datos de referencia. Éstos se pueden resumir en un *Data Warehouse* para ayudar a la accesibilidad y el análisis de los datos.

Como ya se ha mencionado, un *Data Warehouse* es un repositorio para almacenar y analizar información numérica. El propósito de este tipo de repositorios es proporcionar a las empresas o a cualquier entidad interesada en realizar análisis de datos una colección de datos extraídos y desheredados de sistemas operacionales o transaccionales del negocio. El *Data Warehouse* utilizado debe proporcionar soporte en los planes del negocios, además, éste debe ayudar a desarrollar las tácticas que se pueden utilizar para aumentar sus beneficios. No importa la cantidad de información que se tiene, es de poca importancia si no se puede utilizar. La información obtenida puede ser utilizada para el análisis estratégico, investigación de mercados, pronósticos y el análisis de las tendencias.

No obstante que los *Data Warehouse* fueron originalmente creados con el propósito de proporcionar ayuda en la toma de decisiones de las empresas, estos repositorios pueden

ser utilizados para almacenar datos en cualquier otro campo como por ejemplo el académico.

1.3. Minería de datos

La minería de datos puede verse como el resultado de la evolución natural de la tecnología de la información, concretamente en la evolución de las bases de datos en el sentido de la colección de datos y creación de bases de datos, la administración de datos, el análisis y el entendimiento de éstos (Han y Kamber, 2001). En la Figura 1.2 se muestra esta evolución, ahí se puede observar que en primer lugar se desarrollaron las estructuras necesarias para almacenar grandes cantidades de datos, es decir, se crearon las bases de datos. Posteriormente, en los años 70 e inicios de los 80, dada la necesidad de administrar estas grandes cantidades de datos se diseñaron los mecanismos que permitían administrarlos, ejemplos de estos mecanismos fueron las bases de datos jerárquicas, en red, relacionales, el modelo entidad relación, así como lenguajes que permitían realizar consultas sobre estos datos. A partir de mediados de la década de los 80 se han desarrollado herramientas de análisis avanzadas que permiten analizar y entender los datos que están almacenados en estos grandes repositorios de información, como ejemplo de este tipo de herramientas se puede mencionar el proceso de descubrir información en bases de datos. Como se mencionó anteriormente, el surgimiento de la minería de datos está relacionada con dos factores, uno de ellos fue disponibilidad de grandes cantidades de datos almacenados electrónicamente y el otro la necesidad de transformar todos estos datos en información y conocimiento útil que ayudara en la toma de decisiones debido a que la evolución en las bases de datos llevó a una riqueza de datos pero a una pobreza en información. Han y Kamber (2001) definen a la minería de datos como una herramienta para realizar análisis de datos y extraer conocimiento de grandes cantidades de datos, es importante hacer notar que el término se refiere a extraer conocimiento y no a descubrir conocimiento.

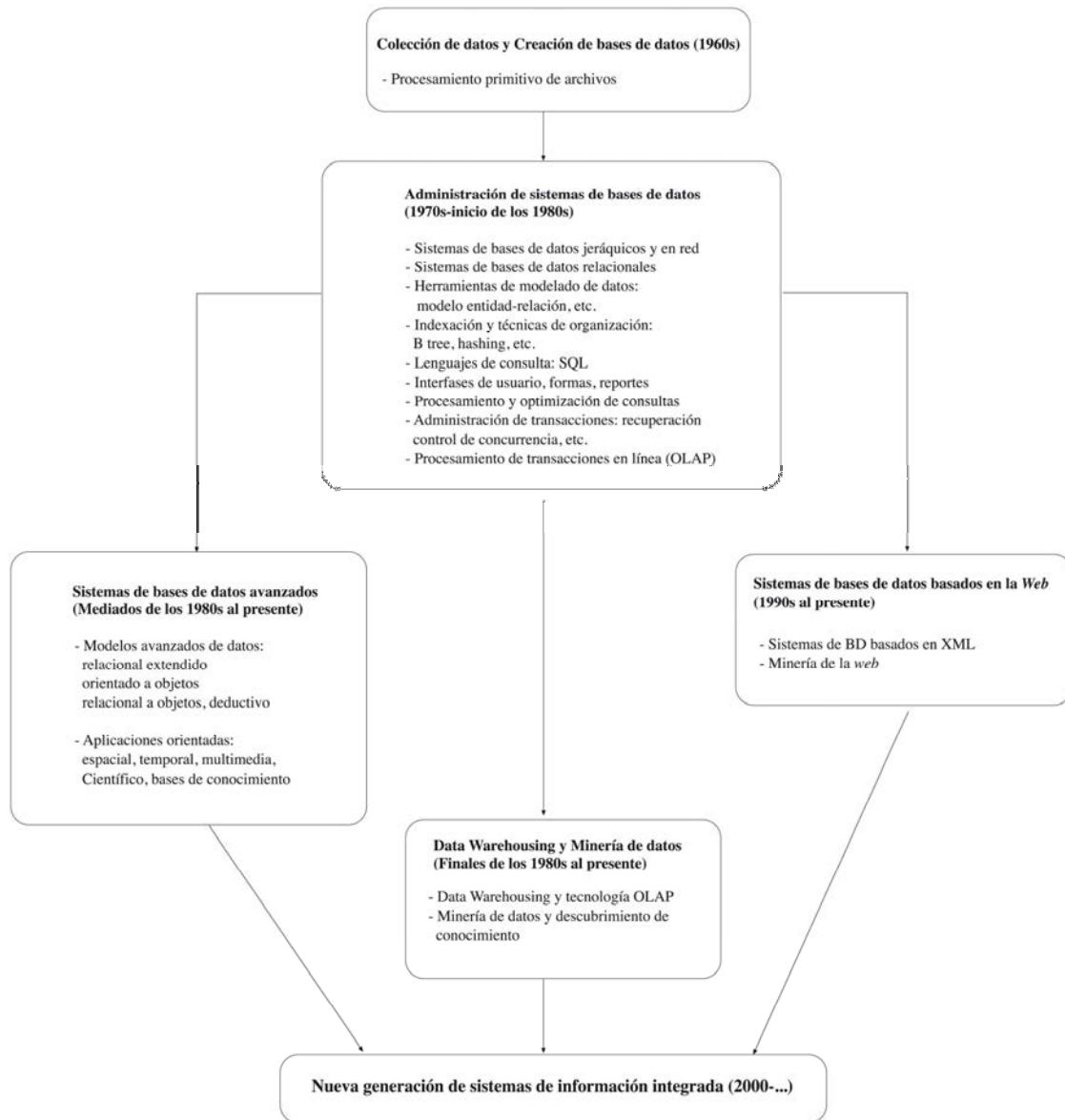


Figura 1.2: Evolución de las tecnologías de Bases de Datos.

Fuente: Han y Kamber, 2001.

La minería de datos se refiere a extraer conocimiento de grandes cantidades de datos y se puede clasificar en dos tipos: i) Minería descriptiva y ii) Minería predictiva. La minería descriptiva describe los conceptos o preguntas relevantes de una forma concisa, breve, informativa y discriminante. La minería predictiva se basa en los datos y en el análisis, construye modelos para la base de datos y pronostica las relaciones y las propiedades desconocidas de los datos. En la Figura 1.3 se muestra una arquitectura típica de un sistema de minería de datos.

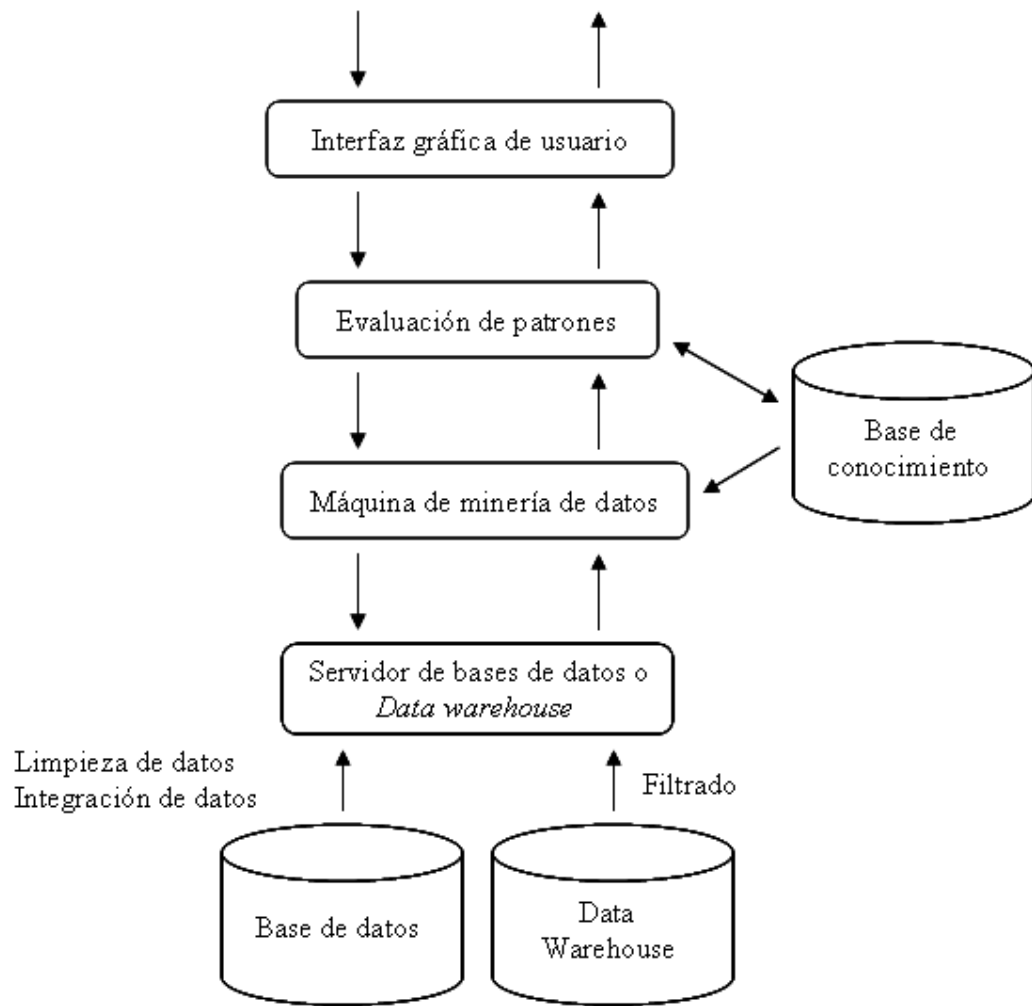


Figura 1.3: Arquitectura de un sistema de minería de datos.

Fuente: Han y Kamber, 2001.

Como se muestra en la Figura 1.4, es importante mencionar que en la minería de datos deben intervenir profesionales de diferentes áreas del conocimiento. Usualmente los profesionales encargados en realizar la implementación de los algoritmos utilizados en la minería de datos son personas relacionadas con las ciencias de la computación, así para que realmente se pueda aprovechar al máximo este tipo de sistemas deben participar expertos en el dominio del problema.

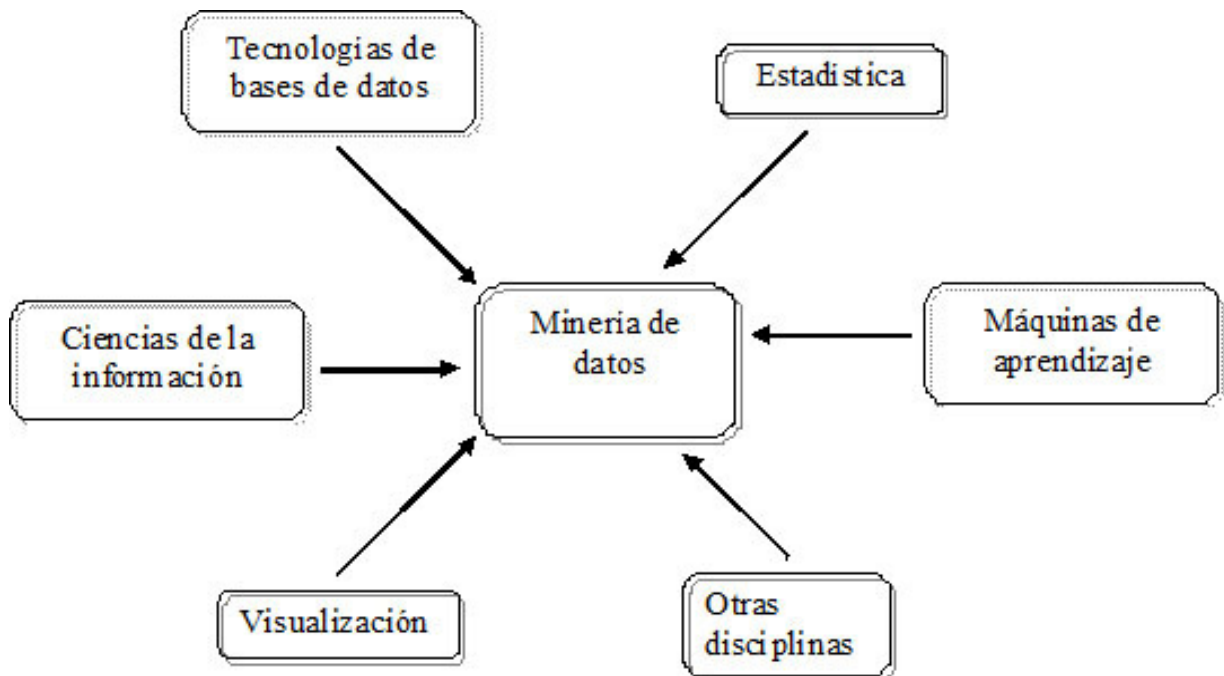


Figura 1.4: Relaciones interdisciplinarias de la Minería de Datos.

Fuente: Han y Kamber, 2001.

1.4. La evaluación de la ciencia

El análisis del desarrollo de la ciencia se inició a principios del siglo XX, al respecto Sancho (1990) asevera que “algunos de los primeros trabajos en bibliometría fueron resultado de una curiosidad innata por entender el desarrollo científico”, pero la necesidad de evaluar la producción científica ha adquirido gran importancia debido a que proporciona información que permite conocer los avances e impacto de la ciencia, muestra el surgimiento de nuevas disciplinas, es posible conocer la forma en que interactúan los países, organizaciones y grupos de trabajo. Además, a partir de ella se pueden asignar recursos económicos en forma más adecuada, así, la actividad científica es vista como una actividad económica.

Las actividades científicas, al igual que el resto de las actividades económicas, para la obtención de sus resultados (bienes de consumo, recursos humanos altamente calificado, información-conocimientos) requiere de determinados recursos (laborales, materiales y de información), es por ello que de la proporción en que estos participan en la actividad o sistema científico, depende en gran medida su nivel de eficiencia

económica (Gorbea-Portal, 2006). Este comportamiento ha motivado el hecho de que el proceso científico pueda ser analizado de forma análoga a los modelos económicos de costo-beneficio o inversión-resultado (Sancho, 1990).

Russell y Rousseau (2002) mencionan que los indicadores utilizados para medir la investigación científica pueden dividirse en dos grandes grupos, estos son: indicadores de entrada (*input*) e indicadores de salida (*output*). Los indicadores de entrada son aquellos relacionados con el dinero invertido, el equipamiento utilizado y el grupo de trabajo. Los indicadores de salida, como los indicadores basados en literatura, representan los efectos y resultados del proceso de investigación. También mencionan que los indicadores pueden ser relativos o absolutos, definen los indicadores absolutos como aquellos que se refieren a una característica en particular de la actividad de investigación, como son número de artículos publicados, número de citas o la cantidad de recursos económicos invertidos, y los indicadores relativos como aquellos que muestran las relación entre dos o más aspectos tales como el número de artículos publicados por grupo de investigación o el número de citas por artículo.

Es importante mencionar que cuando se utilizan indicadores bibliométricos para realizar un análisis, éstos deben ser tomados en conjunto y las fuentes de datos utilizadas deben cubrir adecuadamente la producción científica del grupo temático o disciplina científica que se desea evaluar para que dicho análisis tenga validez. La evaluación de la producción científica cuando se realiza con el propósito de evaluar a un grupo de investigadores tiene mayor validez cuando se realiza en grupos grandes, como por ejemplo análisis de tipo nacional o por disciplina, que en análisis de grupos pequeños, como por ejemplo grupos de investigación, y eso se debe a que los indicadores estadísticos tienen mayor validez cuando se aplican a unidades grandes (Bordons, 2001; Russell y Rousseau, 2002). También la validez de estos tipos de indicadores varía de acuerdo al área que se evalúa ya que las áreas básicas tienen más producción de publicaciones que las áreas tecnológicas o aplicadas.

A continuación se da la definición de algunas técnicas utilizadas en la evaluación de la Ciencia:

1.4.1. Opinión de expertos

Esta técnica se basa en la opinión que proporcionan los expertos en el tema, sobre un trabajo en particular, es posible aplicarlos a uno o más autores. Esta técnica tienen en cuenta las aportaciones de nuevo conocimiento y el nacimiento de nuevas líneas de investigación (Bonilla, 2009). Esta técnica permite determinar la calidad científica de un trabajo en cuanto a su contenido y metodología empleada, tomando en cuenta el contenido científico de cartas, cuestionarios, trabajos publicados, entrevistas acerca de las investigaciones que los científicos llevan a cabo. Otros criterios que son tomados en cuenta son las invitaciones a congresos y conferencias, premios recibidos, honores, nombramientos de sociedades profesionales recibidos por los autores que se están evaluando (Sancho, 1990).

1.4.2. Indicadores bibliométricos

1.4.2.1. De productividad científica

En este tipo de indicadores bibliométricos se tratan aspectos puramente cuantitativos, como son el número y distribución de publicaciones (geográfica, temática, por institución, etcétera), así como la productividad de los autores y su colaboración a través de las publicaciones. Este tipo de indicadores se pueden obtener de distintas fuentes de datos como son las bases de datos bibliográficas, artículos, curriculum vitae, entre otros.

- ***Indicadores por número y distribución de publicaciones***

Los indicadores bibliométricos por número y distribución de publicaciones son los indicadores más comúnmente usados, y son los más sencillos de obtener. Este tipo de indicadores se basan en el número de publicaciones y la distribución que ha realizado un país, una institución, un grupo de investigación, etcétera.

- ***Productividad de los autores***

Estos indicadores muestran el número de publicaciones y su distribución, basándose en la productividad de un autor o grupo de autores en particular.

De acuerdo con Sancho (1990), esta productividad está en función de una serie de variables que pueden agruparse en dos categorías: 1) características personales (inteligencia, perseverancia, capacidad, etc.), y 2) medio ambiente o situación laboral del autor (influencia de colegas prestigiosos, facilidad para obtener información, disciplina en la que está integrado, prestigio de la institución a la que pertenece, dotación económica de la misma, etc.). De manera general, y salvo algunas excepciones, se puede afirmar que existe una fuerte correlación entre la eminencia de un científico y su productividad (Dennis, 1957, citado por Sancho, 1990). La misma autora también menciona que las primeras investigaciones en este campo fueron iniciadas por Lotka en 1926, después continuadas por otros autores, y todos ellos coinciden en que la correlación indicada por Dennis ha existido a través de la historia de la ciencia, y no parece depender del tipo de ciencia o de la fecha elegida. La única condición es que la bibliografía estudiada sea lo más completa posible y cubra un periodo de tiempo suficientemente amplio.

- ***Colaboración en las publicaciones***

Los indicadores de colaboración en las publicaciones son utilizados para conocer la cooperación y actividad que hay entre grupos de científicos, de organizaciones, de países, esto se debe a que el número de artículos producidos por el colectivo es proporcional a su actividad investigadora, y de esta forma se puede considerar como un índice de ella (Sancho, 1990). Existen tres criterios para determinar la colaboración de los firmantes de los documentos, estos criterios son: 1) la publicación se atribuye solo al primer autor; b) se asigna un punto a cada uno de los firmantes; y c) a cada coautor se le puede asignar una fracción proporcional de la publicación. El criterio más comúnmente utilizado es el segundo, esto se debe a que no hay una norma que establezca cual de los firmantes es el primer autor o en que proporción colaboró cada uno de ellos.

1.4.2.2. Indicadores de Conexión o nexos informativos entre trabajos y autores científicos

Este tipo de indicadores se obtiene a través de las notas a pie de página o de las referencias bibliográficas que se agrega al final de las publicaciones para determinar el número de referencias de un trabajo hace hacia otro. Las variables básicas que se pueden utilizar son: el número de referencias por artículo, años de publicación de las fuentes referenciadas, y distribución de las referencias según las revistas o áreas científicas.

1.4.2.3. Indicadores del impacto de los trabajos

El impacto del trabajo de investigación en la literatura juega un papel muy importante porque se refiere al número de citas que un determinado trabajo ha recibido. Este indicador es el más utilizado y también el más polémico, ya que si el número de citas recibidas es muy alto esto no indica necesariamente que el trabajo sea de mayor calidad y viceversa. Algunos investigadores mencionan que ciertos trabajos que han considerado más importantes y de mejor calidad son los que menos citas han recibido.

1.4.2.4. Indicadores de impacto de las revistas

- ***Factor de impacto de las revistas***

Este indicador, introducido por Garfield, muestra la frecuencia con la que un artículo promedio ha sido citado en un año específico. El factor de impacto de una revista de un determinado año se calcula sumando todas las citas recibidas por ésta en ese año, tales citas corresponden a los artículos publicados en los dos años anteriores, y se dividen por el número total de artículos publicados por dicha revista en esos dos años.

- ***Índice de inmediatez***

El índice de inmediatez hace referencia a la rapidez con la que un artículo de una revista determinada es citado; para calcular este índice se toman en

cuenta las citas hechas durante el año en el cual fueron publicados los trabajos citados.

1.4.2.5. Indicadores de visibilidad de las revistas

- ***Índice de influencia de las revistas***

Este índice mide la influencia que tiene una revista (x) en relación con otras revistas (y_n) pertenecientes al mismo grupo, se cálculo a partir de la relación existente entre el número de citas (x) de cada una de las revistas (y_n) y el número de referencias que hace la revista (x) al total de revistas que integran ese grupo (Gorbea-Portal y Suárez-Balseiro, 2007).

- ***Índice de Importancia de las revistas***

Mide la importancia de una revista (x), en relación con otras revistas (y_n) pertenecientes al mismo grupo, se calcula a partir de la proporción existente entre el número de citas que recibe la revista (x) de cada una de las revistas (y_n), y el número total de referencias que hace la revista (x) a todas las revistas, pertenecientes o no al grupo (Gorbea-Portal y Suárez-Balseiro, 2007).

- ***Índice de Posición relativa de las Revistas***

Mide la posición relativa de una revista (x) entre un grupo de revistas (y_n), se calcula a partir de la proporción existente entre el número de citas que recibe la revista (x) de cada una de las revistas (y_n), entre la suma del número total de referencias que hace la revista (x) a las revistas pertenecientes a ese grupo y el total de citas que recibió la revista (x) de todas las revistas del grupo (Gorbea-Portal y Suárez-Balseiro, 2007).

1.5. El estudio bibliométrico de las patentes

Las patentes, unidas a indicadores de ciencia y tecnología tales como los gastos en investigación y desarrollo (I+D), proporcionan una fuente de información con detalles únicos sobre la actividad inventora, es por ello que son de gran importancia para medir la capacidad inventiva de los diferentes sectores productivos de un país. Las patentes

pueden considerarse como un paso intermedio entre la investigación y desarrollo, y la innovación debido a que no todo lo que se patenta representa una innovación. Dada la importancia que juegan las patentes en el desarrollo económico de los países se recurre al estudio de indicadores bibliométricos de patentes.

1.5.1. Patentes

Los indicadores sobre patentes llevan un largo tiempo siendo utilizadas para medir, junto con otro tipo de indicadores, la actividad de ciencia y tecnología (CyT). En el Manual de estadísticas sobre patentes (OCDE, 2009) se menciona que durante los años cincuenta del siglo pasado tuvo un gran impacto el cómputo, realizado por el profesor Jakob Schmookler sobre patentes como indicador del cambio tecnológico en determinados sectores. También Narin (1994) menciona que uno de los primeros trabajos que relacionaban los resultados de la investigación científica y el desarrollo tecnológico fue publicado por Cole y Eales en 1917.

Debido a que las patentes son el objeto de estudio principal de la patentometría, es importante hacer un primer acercamiento a este tipo de documentos antes de iniciar con los indicadores de patentes. Dada la propia naturaleza de las patentes, Griliches (1990) y el Manual de estadísticas sobre patentes (OCDE, 2009) mencionan que las patentes tienen dos tipos de características: i) jurídicas; ii) económicas, las cuales son descritas a continuación.

1.5.2. Fundamentos jurídicos de las patentes

Una patente es un título jurídico otorgado por el estado para proteger una invención (Artículo 28 del Acuerdo de la OMC sobre Aspectos de los Derechos de Propiedad Intelectual Relacionados con el comercio (ADPIC).

1. Una patente conferirá a su titular los derechos exclusivos siguientes:

- a) Cuando la materia de la patente sea un producto, el de impedir que terceros, sin su consentimiento, realicen actos de: fabricación, uso, oferta para la venta, venta o importación para estos fines del producto objeto de patente.

- b) Cuando la materia de la patente sea un procedimiento, el de impedir que terceros, sin su consentimiento, realicen el acto de utilización del procedimiento y los actos de: uso, oferta para la venta, venta o importación para estos fines de, por lo menos, el producto obtenido directamente por medio de dicho procedimiento.

2. Los titulares de patentes tendrán asimismo el derecho de cederlas o transferirlas por sucesión y de realizar contratos de licencia.

Las patentes confieren a los titulares un monopolio de explotación industrial y comercial de las invenciones patentadas durante 20 años (plazo máximo), contados a partir de la fecha de solicitud y al terminar este periodo las invenciones entran en el dominio público. Las patentes confieren derechos territoriales, es decir, si un titular de un invento solamente lo protege en España, el derecho de exclusividad solo será válido en este país y será posible la explotación del invento en cualquier otro país. En este caso la patente es de utilidad al titular ya que el invento no puede ser patentado en ningún otro país debido a que las patentes son concedidas a los *objetos* que son novedosos a nivel mundial.

Una patente puede ser concedida por una oficina de patentes a una empresa, a un individuo o a una entidad pública. Para que un invento sea patentable debe cumplir con tres requisitos esenciales de patentabilidad: a) la invención debe ser una novedad. En el estado de la técnica no debe existir nada similar al invento que se desea proteger a través de una patente, el invento no debió utilizarse con anterioridad, b) debe ser el resultado de una actividad inventiva (de carácter no obvio), es decir, el invento no debe ser evidente para una persona experta en la materia, debe mostrar un avance significativo en relación con el estado de la técnica actual, y c) ser aplicable industrialmente, debe poder utilizarse.

1.5.2.1. Proceso de protección de las patentes

Como se mencionó anteriormente, las patentes pueden ser concedidas por las oficinas de patentes, siendo en estas entidades donde los titulares del invento o inventores deben presentar una solicitud y la oficina se encargará de comprobar que el invento cumpla con todos los requisitos de patentabilidad exigidos. Las solicitudes son examinadas por expertos y es a través de estos exámenes donde se determina si la patente es concedida o denegada.

A continuación se describen las distintas vías que se pueden utilizar para obtener la protección de los inventos:

- **Vía nacional.** Cuando una persona física o moral decide patentar un invento debe presentar su solicitud en una oficina nacional de patentes, usualmente esta oficina nacional es la oficina del país del solicitante. A la primera solicitud de patente que se presenta en el mundo se le denomina solicitud prioritaria, a esta solicitud se le asigna una fecha que también es llamada fecha prioritaria. Después de presentar la solicitud, la oficina de patentes procede a realizar los exámenes necesarios para determinar si el invento cumple con los requerimientos solicitados a los objetos susceptibles a ser patentados. El manual de estadística sobre patentes, entre otros, menciona que el tiempo que transcurre entre la solicitud prioritaria y la publicación de la solicitud es de 18 meses contados a partir de la presentación, a esta segunda fecha se le denomina fecha de publicación de la solicitud. También se menciona que el tiempo transcurrido entre la presentación de la solicitud y la concesión o denegación de la patente oscila entre dos y ocho años; a esta tercera fecha se le llama fecha de concesión.
- **Vía internacional** (Convenio de París). En 1883 se adoptó el Convenio de París para la protección intelectual (WIPO, 2011). Este convenio fue el primer tratado internacional creado para facilitar la protección de creaciones intelectuales (patentes, marcas, y diseños industriales), a 2011 son 173 los países miembro. Gracias a este convenio, si un inventor desea proteger su

invento en más de un país tiene 12 meses, contados a partir de la fecha prioritaria, para presentar la solicitud en cada uno de los países miembro en los cuales desea proteger el invento.

- **Vía internacional** (Tratado de Cooperación en Materia de Patentes). Una alternativa al Convenio de París es el Tratado de Cooperación en Materia de Patentes o PCT (*Patent Cooperation Treaty*). Éste tratado es administrado por la Organización Mundial de la Propiedad Intelectual (OMPI), fue firmado en 1970 y entró en vigor en 1978 (notas del curso de PCT), para el 2011 eran 144 los estados contratantes. Si bien no existe una patente mundial, el PCT facilita el proceso de solicitud de patentes ya que con base en una única solicitud es posible obtener patentes en varios países del mundo.

El proceso que se sigue para solicitar la protección mediante el PCT es el siguiente:

- i) Presentación de la primera solicitud, esta solicitud generalmente se realiza en la oficina nacional de patentes del país de los solicitantes.

Fase Internacional

- ii) Presentación de la solicitud PCT. Tomando como base la solicitud presentada en la oficina nacional se hace la solicitud PCT y se presenta en un plazo máximo de 12 meses contados a partir de la fecha de presentación de la solicitud en la oficina nacional. Ésta se presenta en una oficina receptora que generalmente es la misma oficina nacional (pago único de tasas, un idioma, único conjunto de requisitos formales y efectos jurídicos en todos los estados miembro del PCT). La solicitud PCT tendrá el mismo efecto que una solicitud presentada por separada en cada Estado contratante del PCT.
- iii) Informe de búsqueda internacional y opinión escrita. Aproximadamente 16 meses después de la presentación de la

primera solicitud, el solicitante recibe un informe de búsqueda internacional y una opinión por escrito por parte de la oficina administradora de búsquedas internacionales que le fue asignada. Este informe de búsqueda internacional contiene citas a documentos de patentes y a otro tipo de documentos que estén relacionados con el estado de la técnica del invento, sobre la base de este informe se hace la opinión escrita, la cual es una evaluación preliminar sobre la patentabilidad de la invención.

- iv) Publicación internacional. En teoría, después de transcurridos 18 meses de la fecha prioritaria, la Oficina Internacional de la OMPI hace la publicación internacional de la solicitud de PCT.

Los solicitantes tienen el derecho de ordenar una búsqueda internacional suplementaria y un examen preliminar internacional.

- v) La búsqueda debe solicitarse en cualquier momento, pero antes de que hayan transcurrido 19 meses contados a partir de la fecha prioritaria, tiene como finalidad de mostrar un panorama más amplio del estado de la técnica y debe ser entregado en un plazo de 28 meses, también contabilizados a partir de la fecha de prioridad.
- vi) Examen preliminar internacional. Esta relacionado con el apartado iii), si el solicitante no está de acuerdo con la opinión escrita, puede solicitar un examen preliminar internacional, y tiene como mínimo 22 meses para presentar esta solicitud, que al igual que en los casos anteriores se computa a partir de la fecha prioritaria. Transcurridos aproximadamente 28 meses contados a partir de la fecha prioritaria, la oficina Administradora que realizó este examen debe emitir el informe preliminar internacional sobre la patentabilidad, informe que es

usado por las oficinas de patentes nacionales o regionales para evaluar la patentabilidad de la invención.

Fase nacional

- vii) Tramitación de la solicitud PCT ante las oficinas nacionales o Regionales. Una vez que se ha concluido la fase internacional, aproximadamente 30 meses después de la fecha de prioridad, la solicitud PCT entra en fase nacional, es decir, se inicia con el proceso de solicitud de patentes a escala nacional. Las oficinas de patentes nacionales utilizan los informes de patentabilidad emitidos durante la fase internacional, pero depende de cada una de ellas el otorgamiento de la patente.

La utilización del método de protección vía el PCT ofrece la ventaja de que el solicitante cuenta hasta con 30 meses, contados a partir de la fecha de prioridad, para hacer el pago de los costos relacionados con el proceso de patentamiento, así que ésta es una ventaja sobre el método tradicional al contar con 18 meses más para afrontar los costos. Además, se pueden realizar estudios para determinar en que países realmente se debe proteger el invento, y de ser necesario buscar socios comerciales para llevar a cabo la explotación del invento o sobre la base de los informes de búsqueda, tomar la decisión de continuar o abandonar el proceso de protección. En la Figura 1.5 se muestra gráficamente el proceso de solicitud de patentes utilizando el método (PCT) descrito en este apartado.

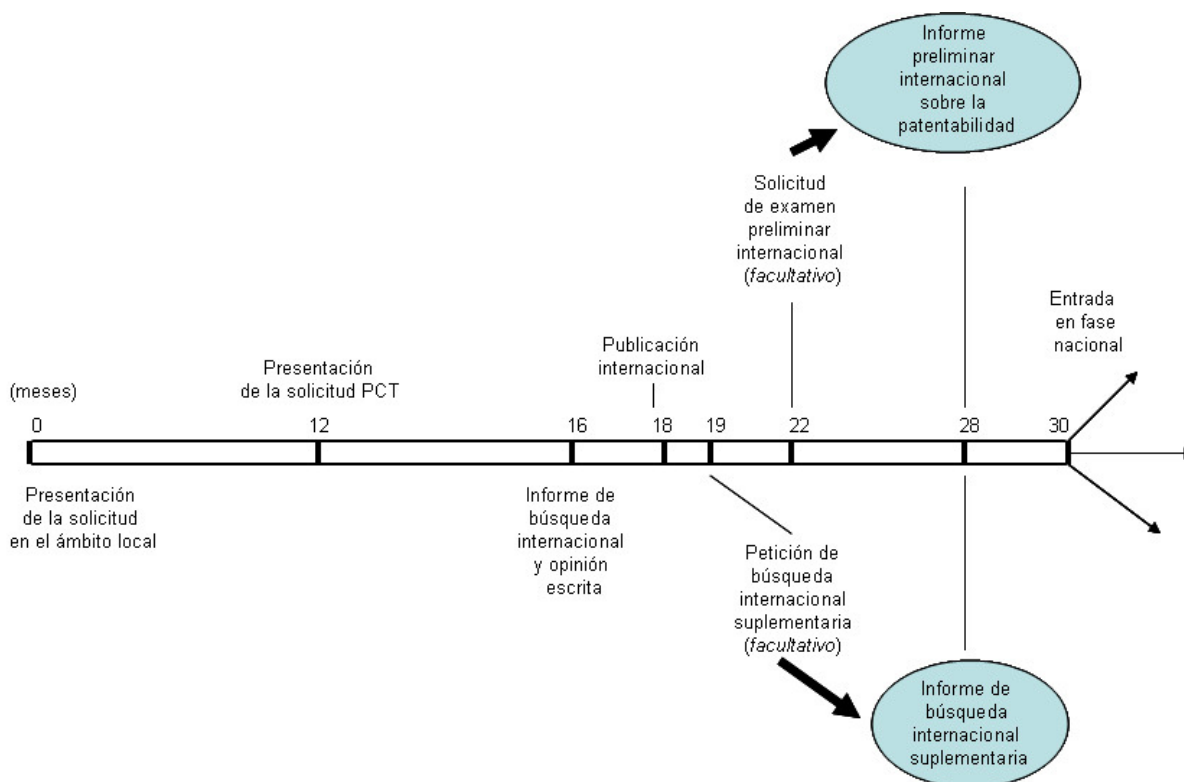


Figura 1.5: Proceso de solicitud de patente mediante el PCT. Fuente: Notas del curso Introducción al PCT

- **Vías regionales.** También existen oficinas regionales como por ejemplo la Organización Africana de la Propiedad Intelectual (OAPI), Organización Regional Africana de la Propiedad Intelectual (ARIPO), Organización Europea de Patentes (OEP), Organización Euroasiática de Patentes (OEAP), en las que los inventores pueden presentar una solicitud de patente.

1.5.3. Fundamentos económicos de las patentes

Las patentes buscan alentar la invención a partir de que ofrecen protección de exclusividad y fomentan su difusión. Es necesario un marco jurídico que las proteja pero que no las haga estáticas, sino que por ejemplo esto sirva como incentivo para los investigadores y para el rol activo que tienen o pueden tener en el mercado. Es preciso aclarar que hay otros métodos además de las patentes para proteger el conocimiento.

La política de patentes debe velar por un equilibrio entre la protección y la difusión. Los costes transaccionales que se deben enfrentar muchas veces amenazan con limitar

o anular la actividad inventiva. Diversas soluciones han surgido para enfrentarlos, como por ejemplo los fondos comunes de patentes y modelos de cámaras de compensación de patentes. Teniendo en cuenta que existe el uso de patentes como instrumento económico, es importante contar con un marco legal que defienda la competencia, que favorezca la difusión y que promueva la protección de invenciones de calidad.

Desde los años ochenta al presente las políticas de patentes se han visto reforzadas como respuesta al papel que han ocupado los avances en el mercado y la economía, y al desarrollo de políticas de I+D. El aumento de la competencia internacional, el progreso científico y tecnológico, entre otro, han promovido la defensa de los derechos de los titulares de las patentes. Datos que muestran estos cambios y el auge de las políticas de patentes son: i) en 1977 la creación de la Oficina Europea de Patentes; ii) en 1982 Estados Unidos crea el tribunal de apelaciones del circuito federal para tratar decisiones en materia de patentes; iii) en 1994 se firma el acuerdo ADPIC/TRIPS (Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio) OCDE, 2009).

1.5.4. Documentos de patentes

Los documentos de patentes son fuentes que contienen información de gran importancia sobre las invenciones, este tipo de información no esta disponible en ninguna parte y por lo tanto son un complemento a las fuentes tradicionales de información utilizadas para medir la I+D+i (OCDE, 2002; OCDE, 2009; Griliches, 1990; Barroso, 2009; entre otros). García-Escudero y López López (1997) dicen que los documentos de patentes responden a dos exigencias legales. Por un lado, deben presentar información técnica suficiente para hacer la divulgación y, por otro, deben delimitar la invención, de tal manera que definan de forma clara el alcance del objeto que se protege, así, la estructura del documento responde directamente a estas exigencias legales.

El documento de patente contiene la información que a continuación se muestra:

- i) Primera página. La página contiene datos bibliográficos como el número de la publicación, número de la solicitud, códigos correspondientes a la clasificación técnica definida por la clasificación internacional de patentes (CIP), fecha de presentación, fecha de publicación de la solicitud, fecha de publicación del folleto de la solicitud, solicitantes, inventores, agente, título de la invención, resumen y un dibujo o fórmula más representativa (en caso de que proceda). En la Figura 1.6 se muestra la primera página de un documento de patente. En este ejemplo se puede observar que hay unos códigos encerrados en un círculo, estos son denominados códigos INID que significa Identificación Numérica Internacionalmente acordada en materia de Datos (bibliográficos), y son administrados por la OMPI. En la Figura 1.6 se observa también los códigos 11 (número de la patente o del documento de patente), 12 (tipo de documento), 19 (código estándar u otra identificación de la oficina de patentes que está haciendo la publicación), 21 (números asignados a las solicitudes), 22 (fechas de presentación de las solicitudes), 43 (fecha de publicación de un documento de patente no examinado), 51 (clasificación internacional de patentes), 54 (título de la invención), 57 (resumen o reivindicaciones), 71 (solicitantes), 72 (inventores), y 74 (agente), en anexo 1 se muestra un listado completo de los códigos INID bibliográficos.

- ii) Descripción. La descripción es lo que se denomina el estado de la técnica. Esta sección debe ser clara y detallada debido a que es en esta parte donde se define el contexto tecnológico de la invención. Se resalta la diferencia de la tecnología anterior y la aportada por la nueva invención. Es la base y fundamento de las reivindicaciones, también es la fuente de información tecnológica más importante y a partir de lo expuesto en la definición cualquier experto en la materia puede poner en practica la invención.

- iii) Reivindicaciones. En las reivindicaciones se define el alcance de la protección que desea hacer el solicitante. Ésta es la parte más importante de los documentos de patentes.
- iv) Dibujos. El o los dibujos solamente se exigen si son necesarios, en éstos se debe describir visualmente la invención y se facilita la comprensión de la misma.
- v) Referencias a patentes y a otros documentos de carácter científico que se consideren de importancia relevante para determinar si se concede o no la patente.

Como se puede observar, el documento de patente contiene tanto características técnicas como historial de la solicitud e información sobre los inventores y solicitantes.



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



⑪ Número de publicación: **2 197 008**
⑲ Número de solicitud: 200201269
⑤① Int. Cl.⁷: A61D 19/02

⑫

SOLICITUD DE PATENTE

A1

⑫② Fecha de presentación: 03.06.2002	⑦① Solicitante/s: MAGAPOR, S.L. Martin Blesa, 31 - 37 50600 Ejea de los Caballeros, Zaragoza, ES
⑫③ Fecha de publicación de la solicitud: 16.12.2003	⑦② Inventor/es: Mena Gericó, Juan José
⑫③ Fecha de publicación del folleto de la solicitud: 16.12.2003	⑦④ Agente: Ungría López, Javier

⑤④ Título: **Catéter de inseminación intrauterina.**

⑤⑦ Resumen:
Catéter de inseminación intrauterina, siendo de utilidad en la inseminación artificial de diversas especies animales y, así, es de aplicación preferente para la inseminación intrauterina de la especie porcina, definiéndose el catéter (1) de inseminación intrauterina por un elemento unitario constituido por un cuerpo tubular (2) alargado dotado en uno de sus extremos, el extremo a introducir en el aparato reproductor del animal, de un cuerpo (3) de superficie redondeada provisto de un orificio (4) como prolongación del orificio (5) axial del cuerpo tubular (2).



Venta de fascículos: Oficina Española de Patentes y Marcas. C/Panamá, 1 - 28036 Madrid

Figura 1.6: Primera página de una patente. Fuente: Base de datos INVENES.

1.5.5. Temas de Investigación sobre las patentes

Debido a la importancia que tienen los indicadores de patentes y a los errores que pueden cometerse en la obtención e interpretación de los mismo; la Organización para la Cooperación y el Desarrollo Económicos (OCDE), en el Manual de Estadísticas sobre Patentes (OCDE, 2009) muestra una serie de temas de investigación en los que se utilizan los indicadores de patentes, así como los indicadores que pueden ser utilizados. Estos temas de investigación se describen a continuación.

1.5.5.1. Rendimiento tecnológico

Los indicadores de patentes se utilizan para medir el rendimiento tecnológico de las empresas, sectores, regiones y países debido a que son indicadores que están más estrechamente vinculados al desarrollo tecnológico. Sobre la base de estos indicadores los administradores de la política científica y tecnológica tienen herramientas de apoyo para identificar las fortalezas y debilidades de los sistemas de innovación nacionales y regionales.

1.5.5.2. Tecnologías emergentes

Con el uso de indicadores de patentes es posible analizar el ascenso de las tecnologías emergentes. Este tipo de análisis es posible gracias a la información detallada que ofrecen los documentos de patentes, y utilizados junto a otros indicadores como los procedentes de publicaciones científicas es posible obtener información más precisa de estas tecnologías.

1.5.5.3. Difusión del conocimiento y dinámicas del cambio tecnológico

Como se ha mencionado, los documentos de patentes contienen, en el apartado estado de la técnica, información detallada del estado de la técnica anterior y del invento que puede ser utilizada para determinar como se transfiere el conocimiento. Además, los documentos contienen citas a patentes y a otro tipo de literatura científica que permiten

realizar análisis para cuantificar el nivel de transferencia del conocimiento entre los diferentes factores participantes en proceso.

1.5.5.4. Geografía de la invención

Dado que los documentos de patentes contienen las direcciones de los inventores y de los solicitantes de las patentes, es posible obtener indicadores que muestren las propiedades geográficas de los procesos inventivos, es decir, qué papel juegan las pequeñas, medianas y grandes empresas, las universidades, etc. en la innovación nacional o a nivel regional, el perfil y la especialización tecnológica.

1.5.5.5. El valor económico de las invenciones

Se ha demostrado que existe una correlación entre el valor de una patente y el número y la calidad de las citas que genera; dado que esta información sobre las patentes esta disponible (me falta poner la referencia), es posible realizar análisis para determinar el valor de una invención y obtener un indicador de importancia económica debido a que el valor de una invención constituye un indicio importante de su repercusión económica.

1.5.5.6. Rendimiento y movilidad de los investigadores

Nuevamente, para obtener indicadores que permitan conocer el rendimiento y movilidad de los investigadores, la dirección es fuente de información de gran importancia. Con este tipo de datos es posible conocer la movilidad de los investigadores entre empresa o países, y encontrar grupos de trabajo.

1.5.5.7. El papel de las universidades en el desarrollo tecnológico

A partir de los datos de las patentes es posible obtener indicadores que son utilizados para determinar el impacto que las universidades han tenido en el desarrollo tecnológico, estos indicadores se obtienen mediante la recopilación de las patentes de las cuales éstas son titulares, también las citas a trabajos de investigación que aparecen

en las patentes juegan un papel muy importante para determinar el impacto de la universidades.

1.5.5.8. Globalización de las actividades de I+D

Los documentos de patentes contienen la nacionalidad tanto de los solicitantes como de los inventores, utilizando estos datos se obtienen indicadores que permiten analizar la actividad y el rendimiento inventivo de las instituciones multinacionales, la titularidad extranjera de invenciones nacionales y viceversa, y medir la intensidad de la co-inventoría nacional.

1.5.5.9. Estrategias de las empresas en materia de patentes

Para identificar las estrategias de mercado de los titulares de las patentes, especialmente los países en los que se busca protección y su orden de importancia, es de gran interés la historia de las solicitudes, ya que a partir de ellas se revela el marco temporal de la invención, es decir, muestra el proceso que ha seguido el titular para patentar el invento y así evidenciar sus estrategias.

Además de los temas de investigación mencionados anteriormente, el Manual de Estadística de Patentes muestra la evaluación de la efectividad del sistema de patentes, previsiones de solicitudes de patentes, y seguimiento del funcionamiento interno del sistema de patentes. Debido a que estas son inherentes a los sistemas de patentes y a las oficinas de patentes no se describen en el presente trabajo.

1.5.6. Indicadores de patentes

La OCDE desde los años setentas había estado trabajando en el campo de las patentes. En 1994 publicó un manual de estadísticas sobre patentes, dada la importancia y auge de los análisis sobre patentes, a través del Foro Consultivo Científico y Tecnológico se dio a la tarea de mejorar y completar este manual para proporcionar directrices básicas que faciliten y normalicen la utilización de indicadores en materia de patentes. Estas

directrices quedan plasmadas en el Manual de estadísticas sobre patentes (OCDE, 2009), a continuación se describen los indicadores propuestos en el manual.

1.5.6.1. Indicadores relacionados con las fechas de referencia

En los documentos de patentes se encuentran la fecha de prioridad, la fecha de solicitud, la fecha de publicación y la fecha de concesión. La fecha de prioridad es la primera fecha en la que se presenta la solicitud de patente en cualquier lugar del mundo, es esta fecha la que se usa para conocer la capacidad inventiva debido a que es la más cercana al momento de la invención.

La fecha de solicitud es la fecha en que se presenta la solicitud en una oficina de patentes determinada. Por lo general el solicitante presenta una solicitud en la oficina nacional, siendo esta fecha la que genera la fecha de prioridad.

La fecha de publicación refleja el momento en que la información relativa a la invención se divulga al público en general y se pone a disposición de los profesionales de la información.

La fecha de concesión es la fecha en la que el organismo autorizado confiere los derechos de patentes al solicitante, la concesión de la patente, dependiendo del organismo que la otorga, en promedio tarda entre tres y cinco años, pero en algunos casos se puede demorar hasta diez años.

1.5.6.2. Indicadores relacionados con el país de referencia

Entre los datos proporcionados por los documentos de patentes se tienen el país de los solicitantes, de los inventores, y de la oficina prioritaria. De aquí que es posible obtener indicadores basados en estos tres tipos de país de referencia.

- El recuento del país de residencia de los solicitantes se puede utilizar para designar la titularidad o el control de las invenciones. Estos indicadores reflejan el rendimiento innovador de las empresas de un país.

- El recuento de patentes relacionados con la residencia de los inventores muestra la capacidad inventiva de los laboratorios locales y de la fuerza laboral de los países.
- El recuento relacionado a las solicitudes prioritarias (oficina prioritaria) es un indicativo del atractivo que representan los procedimientos de patentes de las diferentes oficinas, así como las características económicas generales de los países.

1.5.6.3. Indicadores sobre Familia de patentes

Los indicadores basados en la familia de patentes son de gran importancia para determinar el valor de una invención. Las familias de patentes son usadas en estudios económicos y estadísticos para muchos propósitos, entre ellos se encuentran el análisis de estrategias de patente de los solicitantes, el monitoreo de la globalización de invenciones y la comparación del rendimiento inventivo y el acervo de conocimiento tecnológico de los diferentes países (Martínez, 2011).

Existen varias definiciones de familia de patentes, a continuación se dan algunas de las definiciones de familia encontradas en el Manual de estadísticas sobre patentes (OCDE, 2009):

- Es el conjunto de patentes que protegen una misma invención.
- El conjunto de patentes que se presentan en varios países y se relacionan entre ellas mediante una o varias presentaciones comunes de prioridad.
- Todos los documentos que están vinculados directa o indirectamente a través de uno de prioridad pertenecen a la misma familia de patentes. Esta es la definición que usa el Centro Internacional de Documentación de Patentes o INPADOC (*International Patent Documentation Center*).

- Todos los documentos que tienen como mínimo una prioridad en común pertenecen a la misma familia de patentes. Esta es la definición que se usa en el portal de búsqueda de patentes *espacenet*.
- Todos los documentos que tengan la misma prioridad o prioridades en combinación pertenecen a una misma familia.

Simmons (2009) define la familia de patentes como un grupo de patentes relacionadas con un mismo invento, los documentos de patentes están cumplimentados por el mismo solicitante o grupo de solicitantes en uno o más países. El mismo autor menciona que debido a que las familias de patentes son generadas por los productores de bases de datos y no por las directrices nacionales o internacionales los miembros de las familias de patentes puede variar de una base de datos a otra.

Dada la diversidad y crecimiento de las definiciones de familias de patentes, Martínez (2011) presenta las cuatro definiciones que son más utilizadas, i) patentes equivalentes; ii) familias extendidas, iii) familias basadas en la solicitud prioritaria, y iv) familias validadas por expertos, basadas en el contenido novel de la técnica.

- i) Patentes equivalentes. Estas son consideradas generalmente como la mejor forma para identificar patentes que protegen la misma invención. Son definidas como solicitudes que tienen exactamente la misma prioridad o combinación de prioridades.
- ii) Familias extendidas. Éstas ayudan a encontrar todos los posibles enlaces entre dos patentes dadas. Las familias extendidas se construyen mediante la consolidación de los enlaces prioritarios directos e indirectos entre las solicitudes de patentes en las familias. De esta forma es posible encontrar dos documentos de patentes que no tengan ninguna prioridad en común, pero que estén indirectamente relacionados, ya que ambos comparten al menos una prioridad con una solicitud de terceros.

- iii) Familias basadas en la solicitud prioritaria. Éstas son familias donde cada solicitud prioritaria forma una familia con todos los documentos de patentes presentados posteriormente y que la reconocen como solicitud prioritaria. Una diferencia importante entre este tipo de familia y los otros es que no son mutuamente excluyentes cuando, por ejemplo, dos documentos prioritarios conducen a un documento posterior único. En estos casos cada documento posterior pertenece a dos familias diferentes y por lo tanto se podrá contar dos veces en el recuento total de la familia.

- iv) Familias validadas por expertos basadas en el contenido novel de la técnica. Estas familias difieren de las definiciones anteriores en que el control de expertos se utiliza para validar la composición de la familia de patentes. Se basan en el principio de que una familia debe estar formada por los documentos de patente que protegen el mismo contenido técnico. Los expertos examinan las solicitudes nuevas y certificar si su contenido técnico coincide con el de las familias existentes o no.

Familias de patentes triádicas

Un tipo especial de familia de patentes son las familias de patentes triádicas, este tipo de familias son un conjunto de solicitudes de las mismas presentadas ante la Oficina Europea de Patentes (OEP) y la Oficina Japonesa de Patentes (JPO), y concedidas por la Oficina Americana de Patentes (USPTO), que comparten una o más solicitudes de prioridad.

1.5.6.4. Indicadores sobre la clasificación de las patentes

Los documentos de patentes proporcionan los datos necesarios para clasificarlos en campos específicos del conocimiento y de esta forma obtener indicadores que permiten realizar estudios sobre la detección de nuevos campos técnicos emergentes y su evolución, el ciclo de vida tecnológica (rastrear las tasas anuales de crecimiento de las patentes durante largos periodos de tiempo y así determinar si se produce una reducción de los avances tecnológicos), y conocer la influencia que una tecnología

tiene sobre otra. Para obtener este tipo de indicadores la clasificación más utilizada es la Clasificación Internacional de Patentes (CIP).

1.5.6.5. Indicadores basados en referencias de patentes

El análisis de referencias encontradas en los documentos de patentes ha adquirido gran interés por partes de los profesionales dedicados al estudio de las patentes. El estudio de este tópico tiene sus inicios en los años setenta (Oppenheim, 1997). Los documentos de patentes contienen referencias a patentes y a otro tipo de literatura científica. Éstas se usan para valorar la patentabilidad de la invención y definir la legitimidad de las reivindicaciones de una nueva solicitud de patentes. Las referencias son hechas tanto por los inventores como por los examinadores de las patentes y se utilizan como un indicio del conocimiento científico o tecnológico anterior a la invención, pero también pueden citarse para determinar el grado de novedad y la actividad inventiva que suponen las invenciones.

1.5.6.6. Indicadores del valor de una patente

Las patentes tienen dos tipos de valores determinados en función del punto de vista que se desee analizar, uno es el valor económico y el otro es el valor tecnológico. El valor económico de una patente está relacionado con los ingresos que la patente genera durante su vida útil a el o los titulares de la misma. Este valor está relacionado con la patente en sí, ya que representa la diferencia entre el valor de que la invención está patentada y el que tendría en caso de no estar protegida. El valor tecnológico está estrechamente vinculado al invento, debido a que es medido por las contribuciones que hacen las patentes al acervo del conocimiento tecnológico mundial.

No obstante a que en la actualidad subsiste el debate entre los expertos dedicados a este tipo de estudio, dada la importancia del valor de las patentes, el Manual de Estadística de Patentes de 2009 (OCDE, 2009) presenta algunos indicadores que se pueden utilizar para medir el valor de las mismas. Estos indicadores se mencionan a continuación:

- Recuentos ponderados: ponderar el recuento de las patentes mediante el número de citas posteriores, el número de miembros de la familia, etcétera.
- Recuentos de patentes seleccionadas (desestimación de las patentes de menor valor): familias triádicas, patentes con alto nivel de citas (el 10% más alto de la distribución), de concesiones (en lugar de solicitudes), patentes renovadas hasta una determinada duración (por ejemplo, cinco años), etcétera.
- Número de reivindicaciones en conjunto con las referencias de patentes anteriores, el número de clases técnicas (CIP), número de inventores en una patente, oposiciones y litigios.

Estos indicadores aunado a las grandes cantidades de datos almacenados en las bases de datos bibliográficas de patentes muestran la necesidad de aplicar técnicas que permitan obtener de forma menos compleja la información de la actividad inventiva que esta escondida dentro de estos grandes almacenes de información.

Este marco teórico permite sustentar las necesidades de encontrar estructuras de datos que permitan hacer una explotación de los datos bibliométricos y mostrar las ventajas y desventajas que la aplicación de las tecnologías de la información, ofrece a los Estudios Métricos de la Información. Es por ello que esta tesis tiene como propósito diseñar un modelo teórico-metodológico que presenta la integración de los datos bibliográficos de patentes en un *Data Warehouse*, su implementación en una aplicación *Web*, la obtención de los indicadores y herramientas necesarias para su aplicación en un estudio de caso con las patentes españolas.

CAPÍTULO 2

HIPÓTESIS Y OBJETIVOS

2. Hipótesis y objetivos

Los antecedentes mencionados en el capítulo anterior sobre este tipo de estudio, unido a la importancia que tiene la evaluación de la innovación de los países motivan el interés por identificar patrones de comportamiento de los Datos Bibliométricos de patentes, mediante el uso de los algoritmos utilizados en la Minería de Datos.

2.1. Problema de investigación

¿Cuáles son las relaciones existentes entre los atributos de diversas bases de datos de patentes que permita proponer un Modelo de Integración de éstos en un *Data Warehouse*, con el propósito de mostrar las ventajas que ofrece la organización de los mismos en estos repositorios, así como su explotación a través de las técnicas de la Minería de Datos?

2.2. Hipótesis

Con el fin de proponer una solución al problema antes mencionado, en este trabajo se plantean las siguientes hipótesis:

- Los resultados que se obtienen a través del análisis de las variables utilizadas en los indicadores bibliométricos sobre patentes están directamente condicionados a su vinculación con el sistema bibliográfico del cual se obtienen.
- El análisis de los datos desheredados de los sistemas operativos o bibliográficos y su integración en una estructura tipo *Data Warehouse*, permiten la obtención de nuevos patrones de comportamiento (informativo) no visibles a través del análisis que se obtiene mediante el uso de los indicadores bibliométricos tradicionales de patentes.

- Es factible el diseño de un modelo de integración de datos o variables bibliométricas sobre patentes en una nueva estructura que permita su análisis a través de la Minería de Datos y por consiguiente que facilite la obtención de nuevos patrones de comportamiento (informativo) que se deriven de los datos integrados.

Con el propósito de validar las hipótesis de este trabajo se plantearon los objetivos que a continuación se mencionan.

2.3 Objetivos

2.3.1. Objetivo general

- Proponer un Modelo de Integración de datos Bibliográficos sobre patentes en el que converjan los algoritmos utilizados en la Minería de Datos con los nuevos resultados que se deriven de su aplicación, con el afán de identificar nuevos patrones de comportamiento informativo entre los datos.

2.3.2. Objetivos Específicos

- Estudiar los datos heredados de la estructura bibliográfica, que son utilizados con mayor frecuencia en los indicadores bibliométricos sobre patentes, con el propósito de analizar los sesgos que esta relación provoca en los resultados que se obtienen a través del cálculo de este tipo de indicador.
- Determinar una nueva estructura de datos desheredada del sistema operativo utilizado por las bases de datos bibliográficas de patentes utilizadas como fuente en esta investigación, con la finalidad de obtener un conjunto de datos primarios seleccionados.
- Diseñar una estructura de repositorio de datos tipo *Data Warehouse* que satisfaga las necesidades de los datos primarios seleccionados, así como los

resultados previamente propuestos, con el propósito de facilitar los procesos de la Minería de Datos.

- Comparar los resultados obtenidos mediante el uso de los indicadores bibliométricos sobre patentes y los que resulten de la aplicación del modelo de integración de datos propuestos, con el fin de discutir las fortalezas y debilidades de ambos resultados.

CAPÍTULO 3

DETERMINACIÓN DE FUENTES DE DATOS Y HERRAMIENTAS PARA LA OBTENCIÓN Y PROCESAMIENTO DE INDICADORES DE PATENTES

3. Determinación de fuentes de datos y herramientas para la obtención y procesamiento de indicadores de patentes

En el capítulo 1 se presentó un análisis sobre la importancia que tienen el estudio de las patentes en el ámbito económico y de investigación de los países, y debido a ello resulta imprescindible contar con una herramienta que apoye, a los estudiosos en el área, en la obtención de indicadores de una forma más rápida y fácil. En el presente capítulo se describen las fuentes de datos y las herramientas para su integración y explotación.

3.1. Modelo de integración de datos

3.1.1. Fuente de datos

Como se mencionó en la introducción, un primer paso en el proceso KDD es la selección de las fuentes de datos y su integración en un *Data Warehouse* tomando en cuenta solamente los datos que vayan a ser utilizados; en el presente estudio se seleccionaron tres bases de datos bibliográficos de patentes, que son: a) la base de datos de patentes internacionales *Interpat* del portal de Invenciones en Español INVENES, b) *Derwent Innovations Index*, y c) la base de datos *Worldwide -Full collection of published patent applications from 80+ countries*, que en adelante se denominará *Worldwide-espacenet*.

3.1.1.1. Interpat de INVENES

INVENES es una plataforma de Inventos en Español ofrecida por la Oficina Española de Patentes y Marcas (OEPM), ésta ofrece acceso a las bases de datos *Latipat*, *Diseños*, e *Interpat*.

Latipat es una base de datos bibliográfica de patentes y modelos de utilidad de dieciocho países iberoamericanos con una cobertura desde 1995, también contiene sus imágenes desde 1991 y documentos completos mexicanos. La base de datos *Diseños* contiene datos bibliográficos de modelos, dibujos y diseños industriales, su cobertura

es a partir de 1966. Incluye las imágenes publicadas en el Boletín Oficial de la Propiedad Industrial (BOPI) desde 1998. *Interpat* es una base de datos que contiene datos bibliográficos de Privilegios Reales desde 1826 hasta 1878, documentos de patentes y modelos de utilidad tramitados por el Estatuto de la Propiedad Industrial desde 1929 y por la nueva Ley de Patentes de 20 de marzo de 1986, así como las patentes europeas y patentes solicitadas vía PCT que designan a España y generan un documento en español. Incluye las imágenes publicadas en el BOPI desde 1988, así como los documentos completos. Estas características, unido a la necesidad de diseñar un modelo de *Data Warehouse* para el estudio de las patentes españolas como parte de esta investigación de tesis, constituyeron las razones principales por las que se decidió utilizar la base de datos *Interpat* en este estudio, además debido a que este tipo de base de datos nacionales proporcionan datos más específicos y actualizados que las bases de datos internacionales.

Interpat proporciona una interfaz de búsqueda básica, avanzada, y experta. El formulario de búsqueda avanzada incluye campos de búsqueda prediseñados, incluye búsquedas en el título, resumen, número de publicación, número prioritario, fechas de publicación, nombre del solicitante, nombre del inventor, o clasificación (ver Figura 3.1).

El formulario de búsqueda de expertos (ver Figura 3.2) es similar a una interfaz de línea de comandos, y permite, mediante el uso de operadores booleanos, buscar a través de varios campos en una sola consulta. Además de esta línea de comandos, la interfaz de búsqueda de expertos permite acceder a un formulario de búsqueda que ayuda a seleccionar los campos y operadores booleanos, la definición de las etiquetas utilizadas en la búsqueda esta disponible a través de del botón de ayuda (i).

Figura 3.1: Pantalla de búsqueda avanzada en la interfaz *Interpat*.

Figura 3.2: Pantalla de búsqueda experta en la interfaz *Interpat*.

Una vez que se ha realizado la búsqueda, esta interfaz cuenta con una opción que permite descargar los registros, los cuales son guardados en un archivo en formato del procesador de palabras *Word*. Las etiquetas mediante las cuales se diferencian los datos de los registros bibliográficos descargados se muestran en la Tabla 3.1.

Tabla 3.1: Lista de etiquetas correspondientes a la base de datos *Interpat*.

Etiquetas
Categoría de los documentos nombrados en el IET
Clasificación de invenciones (8a en adelante)
Clasificación adicional indexada (8 en adelante)
Clasificación anterior a la octava edición
Clasificación de invenciones anteriores a octava edición
Clasificación europea
Código de la provincia del solicitante
Código de nacionalidad del inventor
Código de países designados en EPO
Código de publicación de la patente española
Dirección del solicitante
Fecha de concesión de la patente europea
Fecha de concesión de la solicitud de patente nacional
Fecha de la solicitud de la patente europea (en EPO)
Fecha de publicación del documento A4/T1
Fecha de publicación del informe sobre el Estado de la Técnica
Fecha de publicación de la solicitud europea (en EPO)
Fecha de publicación de la solicitud nacional (en OEPM)
Fecha de publicación del IET (en EPO)
Fecha de publicación en BOPI de la concesión nacional (en OEPM)
Fecha de solicitud de la patente nacional (en OEPM)
Fecha del Tomo IV
Indexaciones a la clasificación anteriores a la octava edición
Informe sobre el estado de la técnica de la OEPM (documentos citados en el IET)
Inventores
Nacionalidad del solicitante
Número de publicación de la solicitud europea (en EPO)
Número de publicación de la solicitud nacional (en OEPM)
Número de solicitud europea (EPO)
Número de solicitud nacional (en OEPM)

Referencia
Resumen
Solicitantes
Tipo de documento
Tipo de patente
Tipo de publicación europea
Título

Fuente: Elaboración propia.

3.1.1.2. *Derwent Innovation Index*

La base de datos *Derwent Innovation Index* es un producto elaborado por *Thomson Reuters*. *Derwent Innovation Index* es una herramienta que combina *Derwent world Patents Index*, *Patents Citation Index*, y *Derwent Chemistry Resources*. Esta base de datos almacena más de 11 millones de inventos básicos y 22 millones de patentes, con cobertura desde 1963, cada semana se le añaden aproximadamente 25,000 registros nuevos de patentes. La información de las patentes se toma de 40 autoridades emisoras de patentes de todo el mundo y se clasifica en tres categorías o secciones: Química, Ingeniería, y Electricidad y Electrónica. Su banco de referencia, desde 1973, son 6 autoridades emisoras de patentes de gran importancia como son el Tratado de Cooperación en Materia de Patentes, Estados Unidos, Europa, Alemania, Gran Bretaña y Japón (Thomson C, 2004).

Como ya se mencionó, *Derwent Innovation Index* es un derivado de la base de datos *Derwent World Patent Index*. Martínez (2010) menciona que las familias de patentes, en la base de datos *Derwent World Patent Index*, son construidas tomando como base el principio de novedad, lo cual implica que para agregar un nuevo miembro a la familia es necesario que el contenido técnico de esta nueva patente sea equiparable con el contenido técnico del resto de miembros de la familia. Es así que las familias de patentes de *Derwent World Patent Index* toman en cuenta tres criterios: i) para identificar a los miembros de una familia de patentes se consideran los vínculos entre las solicitudes prioritarias; ii) se crea una nueva familia de patentes cuando se estima

que una solicitud tiene una contribución básica; y iii) el control de expertos para validar el contenido técnico equiparable y construir la familia.

Por lo expuesto aquí, se podría decir que las familias de patentes proporcionadas por la base de de datos *Derwent Innovation Index* son familias validadas por expertos.

Derwent Innovation Index proporciona una interfaz de consulta (ver Figura 3.3) que permite obtener los datos de las patentes desde 1963 a la fecha (la licencia contratada por la Fundación Española para la Ciencia y la Tecnología (FECYT) solo permite obtener datos de 1980 a la fecha). A través de esta interfaz se pueden realizar búsquedas por tema, título, inventor, solicitante, número de patente, clasificación internacional de las patentes, empresa solicitante, códigos asignados por *Derwent*, y número de identificación único asignado por *Derwent Innovation Index*, entre otros.

The screenshot displays the Derwent Innovation Index search interface. At the top, there are navigation links like 'Go to mobile site', 'Sign In', 'Marked List (0)', 'My EndNote Web', 'My ResearcherID', 'My Citation Alerts', 'My Saved Searches', 'Log Out', and 'Help'. Below this, there's a 'Select a Database' section with 'Derwent Innovations Index' selected. The main search area has a 'Search' button and a 'Clear' button. It includes a 'Current Limits' section with 'Timespan' set to 'All Years' and 'Citation Databases' set to 'Chemical Section; Electrical and Electronic Section; Engineering Section'. The right sidebar contains a 'FECYT' logo and a 'Questions about the new Web of Knowledge?' section with links to 'Have questions about differences in citation counts...', 'Looking for Book Citation Index?', and 'Support, Tools, Tips Training & Support'.

Figura 3.3: Pantalla de búsqueda en la interfaz de *Derwent Innovation Index*.

Al igual que en los otros productos *Thomson Reuters*, la interfaz de *Derwent Innovation Index* proporciona los formatos *Save to other references software*, *Save to HTML*, *Save to plain text*, *Save to tab delimited (Win)*, y *Save to tab-delimited (Mac)* para descargar los registros bibliográficos seleccionados; en el presente trabajo se utilizó el formato denominado *Plain Text* por ser el que más facilita el proceso de migración de datos a una base de datos. Este formato permite guardar los registros bibliográficos de patentes en formato de texto plano con etiquetas que diferencian a cada uno de los datos bibliográficos, es de gran importancia mencionar que cada uno

de los registros bibliográficos de patentes proporcionados por *Derwent Innovation Index* describen una familia de patentes. En la Tabla 3.2 se describen las etiquetas de los datos proporcionados con el formato *Plain Text*, mismos que serán utilizados para generar una primera aproximación del esquema del *Data Warehouse*.

Tabla 3.2: Lista de etiquetas correspondientes a la base de datos Derwent.

Etiquetas	Descripción
PT	<i>Tipo de documento.</i>
GA	<i>Identificador único</i> , también asignado por <i>Derwent Innovation Index</i> . Los valores de este identificador son iguales a los valores contenidos por el atributo UT, la diferencia entre ellos radica en el formato. Este identificador se corresponde con el año de publicación de la primera patente de la familia de patentes del registro, seguido por 6 dígitos. Estos dos valores están separados por dos guiones (–) del resto de caracteres.
AU	<i>Nombre de los inventores</i> en formato apellido iniciales. El apellido puede contener un máximo de 30 caracteres seguido por un espacio y hasta 3 iniciales, si es más de un inventor los nombres están separados por el carácter (,).
AE	<i>Nombre de las empresas o individuos solicitantes y códigos</i> . Estos datos corresponden a las empresas o individuos a quienes se les han transferidos los derechos de titularidad de la patente; es importante señalar que en algunos casos los inventores y los solicitantes pueden ser los mismos.
TI	<i>Título de la patente</i> . Es una descripción concisa en inglés del título, es escrita por analista de <i>Derwent Innovation Index</i> para remarcar el contenido y novedad de la invención descritas en las especificaciones de la patente.

PN	Listado de todos los números de patente de todos los miembros de la familia de patentes. El número de patente es un número serial asignado a cada documento por las oficinas de patentes. <i>Derwent Innovation Index</i> muestra los dos caracteres del código del país al que corresponde la patente, seguido por el número serial (hasta 10 dígitos), y el código del estatus de la patente, éste código indica el tipo de documento o la etapa de publicación. Los números de patentes están separados por (;). ejemplo de número de patente: WO200079505-A1. Este atributo es el equivalente al atributo NUM PUB OEPM (npp) de la base de datos española y es mediante el cual se relaciona con los datos extraídos de la base de datos <i>Derwent Innovation Index</i> .
AB	Resumen. Sobre la base de las divulgaciones y reivindicaciones de las patentes, los analistas de <i>Derwent Innovation Index</i> hacen un resumen conciso que cubre las principales reivindicaciones de la patente. Es presentado por categorías como son la novedad, descripción detallada, uso, ventajas, entre otras. Las categorías presentadas dependerán de la patente seleccionada.
EA	Resumen equivalente. Este resumen corresponde a patentes que comparten el mismo invento y la misma fecha de prioridad, pero fueron emitidas por una autoridad diferente. Este resumen también es preparado por analistas <i>Derwent Innovation Index</i> y se divide en las mismas categorías que el resumen.
DC	Códigos Derwent. Éste es un sistema de clasificación <i>Derwent Innovation Index</i> aplicado a todas las patentes. Todas las tecnologías están cubiertas por 20 áreas, cada una de las áreas están subdividida en clases. La clasificación correspondiente a Química e Ingeniería es realizada por especialistas de <i>Derwent Innovation Index</i> , y las clasificaciones correspondientes a Electricidad y Electrónica se hacen automáticamente de acuerdo a la clasificación Internacional de Patentes.

MC	Códigos manuales Derwent. Estos códigos son asignados por indexadores de <i>Derwent Innovation Index</i> , son utilizados para indicar la novedad en los aspectos técnicos de una invención, así como de sus aplicaciones.
IP	Clasificación internacional de patentes (CIP). Esta clasificación es un sistema reconocido internacionalmente y es controlado por la Organización Mundial de la Propiedad Intelectual (OMPI). Los códigos son asignados por las oficinas de patentes que publican el documento de patente.
PD	Familia de patentes. Aquí se encuentran datos de cada uno de los miembros de la familia de patentes del registro, los datos mostrados son: el número de la publicación, la fecha de publicación, la clasificación internacional de patentes principal, la semana en la que el registro fue ingresado a la base de datos, número de páginas, y el idioma original de la patente.
AD	Detalles de solicitud y fecha. En este atributo encontramos el número de patente asignado por la oficina de patentes local y la fecha. La fecha de solicitud es la fecha en la cual se hizo la solicitud en la oficina local.
FD	Otras detalles de la solicitud. En esta parte se encuentra más información acerca de la solicitud. Como por ejemplo la información relacionada con tres tipos de solicitudes de patentes relacionadas con la patente prioritaria (continuación, continuación en parte, y divisionales). Las solicitudes de tipo continuación se refiere a la segunda solicitud o solicitudes subsecuentes que son hechas mientras que la solicitud prioritaria está pendiente; estas solicitudes reivindican el mismo invento de la solicitud prioritaria pero se benefician de la fecha de solicitud de ésta (Patente Lens, 2012 y Boot Udall 2010). Las solicitudes de tipo continuación-en-parte pueden ser utilizadas para ampliar la cobertura de solicitudes anteriores para incluir mejoras y

	<p>modificaciones de unan invención, para solicitar reivindicaciones que fueron desestimadas en la solicitud principal, entre otras cosas (Patente Lens, 2012 y Boot Udall 2010).</p> <p>Las solicitudes de patentes de tipo divisional son hechas cuando se determina que la solicitud de patentes prioritaria es compleja y debe ser dividida en varias solicitudes conservando la fecha de prioridad de la solicitud original. Los dos primeros tipos de solicitudes es común encantarlas solo en EE.UU., mientras que las solicitudes del tipo divisional es posible encontrarlas es otros países ya que así lo exige el Convenio de París en el artículo 4G.</p>
PI	<p>Información de la solicitud prioritaria y fecha. Bajo los derechos del Convenio de París y del PCT, las patentes pueden solicitarse en uno o más de los países contratantes a los 12 ó 30 meses (según la vía que se haya elegido) contados a partir de la primera solicitud. Así, esta primera solicitud y fecha se convierten en la solicitud y fecha prioritaria, son estos datos los que se encuentran en este campo.</p>
DS	<p>Estados designados. En esta parte del registro se encuentran los códigos de los países en los que se solicitó patentar el invento. Los códigos de estos países se muestran en dos categorías: a) Nacional, indica los estados designados en los que la invención se protegió mediante una patente nacional, y b) Regional, indica los estados en los que se solicitó la protección a través de oficinas de patentes regionales.</p>
FS	<p>Campos de búsqueda. El campo de búsqueda se utilizan las clases internacionales y de EE.UU. para indicar el área temática general de que el examinador de la Oficina de Patentes ha buscado para asegurar que la invención fuera novedoso e involucró una actividad inventiva.</p>
CP	<p>Patentes referenciadas. En este atributo se encuentran las patentes citadas por los inventores y/o por los examinadores de la familia de patentes cuyos números de patente aparecen en PN.</p>

CR	Otro tipo de documentos referenciados. En esta sección se tiene una lista de otro tipo de documentos citados por el inventor y/o citados por el examinador. Al igual que en el caso de las patentes citadas, los documentos citados aquí corresponden a la familia de patentes mostrado en PN.
DN*	<i>El número de recursos de química Derwent (Derwent Chemistry Resources Number)</i> es un identificador de registro único que indica cuando dos o más compuestos están relacionados entre sí. Consisten en un sistema de identificadores, donde la primera parte del código es el mismo para todos los compuestos relacionados, y los sufijos indican ciertas características particulares.
MN*	<i>Números Markush.</i> Un término <i>Markush</i> denota una sustancia o sustituyente, agente reactivo u otro material que se describe como parte de un grupo que consta de ciertos materiales específicos. El valor de las estructuras <i>Markush</i> en Patentes es que diferentes compuestos químicos se pueden describir en una sola reivindicación de la patente; estas estructuras se permiten en las patentes para proteger, en una invención, los compuestos relacionados sin necesidad de que el inventor tenga que preparar y ponga a prueba cada uno compuesto posible (Austin, 2001).
RI*	<i>RIN.</i> Muestra los números de llamada índice asociados con el registro de patentes. Números de llamada Index (RIN) son códigos de cinco dígitos asignados a los sistemas de anillos químicos que no están definidos con precisión por códigos químicos de fragmentación (M0-M6) de <i>Derwent World Patents Index</i> (DWPI). Estos se pueden buscar en DWPI a partir de 1972, para las patentes clasificadas en las secciones B (productos farmacéuticos), C (Agroquímicos) y/o D (Productos Químicos en general). Puede haber cero o más RIN para cada registro de patentes.

UT	Identificador único , asignado por <i>Derwent Innovation Index</i> , para cada registro. Estos valores inician con las siglas DIIDW, seguidas del carácter (:).
*	Las etiquetas marcadas con este símbolo pertenecen a la sección de química

Fuente: Elaboración propia.

3.1.1.3. El Servicio europeo *esp@cenet*

La Organización Europea de Patentes es una organización intergubernamental creada el 7 octubre de 1977 bajo la Convención Europea de Patentes firmada en 1973. Esta organización está formada por dos entidades: i) la Oficina Europea de Patentes (EPO, por sus siglas en inglés), y ii) el Consejo Administrativo que supervisa las actividades de la Oficina.

La Oficina Europea de Patentes ofrece a los inventores un proceso de aplicación uniforme que les permite solicitud protección a través de patentes en los estados miembros de la EPO. A la fecha son 38 miembros los estados contratantes y dos en los cuales se reconoce la patente europea a petición (EPO, 2012). En el anexo se muestra el listado completo de los países contratantes de la EPO.

La divulgación de las invenciones y publicaciones posteriores son fundamentales para el sistema europeo de patentes. El público puede obtener copia oficiales de documentos de patentes europeas a través del Servidor de la Publicación Europea. El Registro de Patentes Europeas ofrece información sobre el estado procesal de las solicitudes de patentes en la OEP. La EPO cuenta con una colección de más de 70 millones de documentos de patentes de todo el mundo, y está a disposición del público, de forma gratuita en Internet, a través del servicio *Espacenet*.

La EPO, a través de su portal de búsqueda de patentes *Espacenet*, ofrece acceso a las bases de datos *Worldwide-espacenet*, *EP-espacenet*, y *WIPO-espacenet*. La base de datos *Worldwide* ofrece acceso a una colección de publicación de patentes de más de

80 países, *ES-espacenet* cuenta con una colección de publicación de patentes europeas que incluyen texto completo, y la base de datos *WIPO-espacenet* ofrece acceso a una colección de patentes publicadas del PCT, también a texto completo.

Para el presente trabajo se uso la base de datos *Worldwide-espacenet*, esta base cuenta con una interfaz de búsqueda mediante la cual se puede buscar a través de palabras clave en el título, palabras clave en el título o resumen, número de publicación, número de solicitud, número prioritario, fecha de publicación, solicitantes, inventores, clasificación europea de patentes, y clasificación internacional de patentes (Figura 3.4).

The screenshot shows the Espacenet Patent search interface. The header includes the Espacenet logo and navigation links for Deutsch, English, Français, and Contact. The main content area is divided into sections for search options and advanced search criteria.

Search Options:

- Smart search
- Quick search
- Advanced search**
- Number search
- Classification search

Advanced search

1. Database

Select the database in which you wish to search:

2. Search terms

Enter keywords in English - ctrl-enter expands the field you are in

Keyword(s) in title:

Keyword(s) in title or abstract:

Publication number:

Application number:

Priority number:

Publication date:

Applicant(s):

Figura 3.4: Pantalla de búsqueda de la interfaz espacenet

Esta interfaz también proporciona una opción que permite descargar los registros en archivos con formato *CSV* (del inglés *comma-separated values*). Este formato se utiliza para convertir datos entre diferentes programas de hojas de cálculo, los valores se almacenan en texto plano, es decir, sin ningún tipo de codificación, cada uno de los valores es separado con una coma, y el registro lo finaliza con un cambio de línea. En

el Tabla 3.3 se muestra una lista de los atributos de los registros que permite descargar esta herramienta.

Tabla 3.3: Lista de etiquetas presentes en los registros de la base de datos *Worldwide-espacenet*.

Etiquetas
Título
Número de publicación
Fecha de publicación
Inventores
Solicitantes
Clasificación Internacional de Patentes
Clasificación Europea de Patentes
Número de solicitud
Fecha de solicitud
Números prioritarios
Patentes citadas en el reporte de búsqueda
Literatura citada en el reporte de búsqueda
Patentes citadas por el examinador
Literatura citada por el examinador
Otras patentes citadas
Otra literatura citada
Patentes usadas en oposición
Literatura citada en oposición
Patentes citadas por el solicitante
Literatura citada por el solicitante

Fuente: Elaboración propia.

3.1.2. Modelo de *Data Warehouse*

Después de hacer la selección de fuentes de datos, como son las bases de datos y/o archivos planos, otro paso importante en el modelado de los Repositorios de Datos o *Data Warehouse* es la selección de los atributos que serán tomados en cuenta en el

modelo. Para la selección de esos atributos se tomarán como base los indicadores propuestos en el Manual de estadísticas de patentes (OCDE, 2009).

Una vez que se han seleccionado los atributos de estas tres bases de datos se diseñará el modelo de *Data Warehouse* utilizando la metodología de bases de datos relacionales. Para su explotación, este diseño se creará en el administrador de bases de datos relacional MySQL. En este punto es importante mencionar que, como era de esperarse, el atributo que tienen en común las bases de datos *Derwent Innovation Index*, *Inerpat*, y *Worldwide-espacenet* es el número de patente, de tal manera que es éste el atributo que se utiliza para enlazarlas. Así mismo, se detectó que no obstante que la base de datos *Worldwide-espacenet* muestra los atributos Literatura citada en el reporte de búsqueda, Patentes citadas por el examinador, Literatura citada por el examinador, Otras patentes citadas, Otra literatura citada, Patentes usadas en oposición, Literatura citada en oposición, Patentes citadas por el solicitante, y Literatura citada los valores para cada uno de los registros buscados no muestran estos datos.

MySQL es un administrador de bases de datos relacionales disponible en diferentes plataformas, entre ellas se encuentran *Microsoft Windows*, *Linux*, *Unix*, y *Mac OS X*, este ostenta la GNU *General Public Licence* (MySQL, 2012), lo cual significa que es un software libre, este tipo de licencias pretenden garantizar la libertad de compartir y modificar todas las versiones de un *software* (GNU, 2012).

La herramienta utilizada para el modelo de datos es el modelador de datos *CASE Erwin*, con ésta se obtienen diagramas de datos claros y de fácil interpretación. La notación presente en los modelos diseñados con *Erwin* es la siguiente:

- **Relaciones.** Las relaciones son representadas por rectángulos que contienen en la parte superior los atributos que forman parte de la llave primaria y en la parte inferior el resto de atributos. Un rectángulo "redondeado" es una relación hija, mientras que un rectángulo con esquinas es indicativo de una relación padre.

- **Interrelación.** Las interrelaciones entre relaciones se modelan mediante una línea continua que indica que los atributos que conforman la llave primaria de la relación padre son llaves foráneas en la relación hija y forman parte de la llave primaria de esta última. Si la línea es punteada es indicativo de que la llave foránea es un atributo más en la relación hija.
- **Cardinalidad.** La cardinalidad se muestra mediante el uso de un punto negro en las interrelaciones. Si no hay puntos negros en los extremos de las interrelaciones significa que la interrelación entre las dos relaciones es uno a uno, si hay un punto negro en la relación hija significa que un registro de la relación padre puede tener uno o más registros en la relación hija.

3.2. Tratamiento de los datos

Otro paso importante en el proceso KDD es el tratamiento de los datos almacenados en el *Data Warehouse* construido, en esta sección se describe el procedimiento utilizado para realizar el tratamiento de ellos. Se ha utilizado el lenguaje de programación Perl para implementar *scripts CGI (Common Gateway Interfases)*, el *DBD::mysql (DataBase Driver)* y el *DBI-mysql (DataBase Interfase)* para realizar la conexión entre el lenguaje de programación y el administrador de bases de datos MySQL. También utiliza un servidor HTTP Apache, instalado en un servidor con sistema operativo UNIX.

El lenguaje de programación *Perl*, desarrollado originalmente por *Larry Wall* y utilizado en sus inicios por administradores de sistemas, es ahora un lenguaje bien establecido que se ha ido desarrollando a través del tiempo y el esfuerzo de un sinnúmero de programadores de *software* libre, es una herramienta muy poderosa que puede ser utilizada en prácticamente todos los sistemas operativos y con el tiempo se ha convertido en un lenguaje sumamente robusto utilizado por los desarrolladores *Web* y programadores de todo el mundo (Wall, 1996; Cozens, 2000).

Un CGI es un protocolo que especifica la manera en cómo los scripts escritos por los programadores, que son ejecutados en un servidor *Web*, pueden comunicarse con el

navegador (cliente). Los *scripts* CGI son pequeños programas que siguen el protocolo CGI, éstos actúan como interfase entre un usuario *Web* (cliente) y el servidor *Web*, estos programas son ejecutados por el servidor *Web*. Los *scripts* CGI aparecen como enlaces dentro de una página HTML y pueden recibir datos de entrada tomados de formas HTML, los resultados pueden ser texto y/o gráficos. La forma de operar un *script* CGI es la siguiente: 1) el script es invocado por el servidor *Web*, 2) lee y decodifica las entradas de los usuarios, las cuales fueron enviadas típicamente a través de formas HTML, 3) realiza el procesamiento y si es necesario hace una conexión a una base de datos, 4) genera dinámicamente una página *Web*, y 5) la página es enviada al navegador (cliente) a través del servidor *Web* (Figura 3.5).

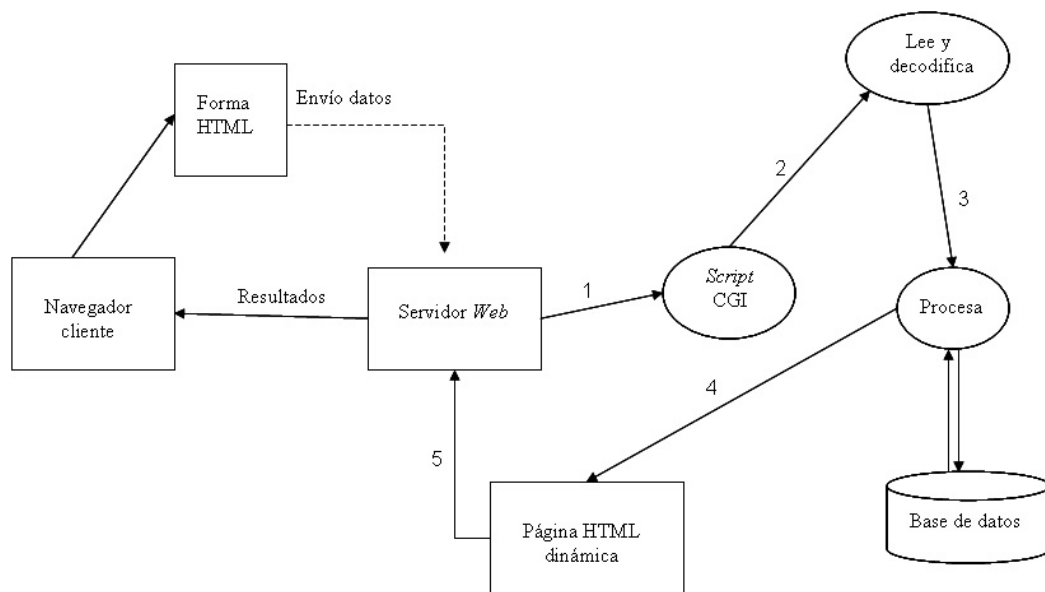


Figura 3.5: Funcionamiento de los scripts CGI

El sistema aquí desarrollado no sigue la metodología orientado a objetos, pero se usó el Lenguaje del Método Unificado (UML, por sus siglas en inglés) y los diagramas de secuencia de ésta debido a que muestran de una forma clara la secuencia de los *scripts* implementados y la herramienta para hacer el modelado de estos diagramas es *starUML*. Los diagramas de secuencia expresan las interacciones entre los actores del sistema, haciendo énfasis especial en el tiempo. Se pueden identificar objetos, responsabilidades, atributos e interrelaciones entre las relaciones.

Con las herramientas arriba mencionadas se implementaron los *scripts* para llevar a cabo las actividades de poblado del *Data Warehouse*, así como la normalización de datos, y generación de reportes para su posterior utilización en herramientas de análisis y visualización.

3.3. Visualización de datos

Otro paso de gran importancia en la metodología KDD es la visualización ya que éste es un proceso útil para aprovechar las capacidades perceptivas de los seres humanos para encontrar las características de estructura de la red y los datos. Para este proceso se utilizaron las herramientas de visualización descritas en los siguientes epígrafes. Las herramientas de visualización *Gephi* y *VOSviewer* tienen su formato de datos propios pero también aceptan el formato de entrada de los archivos usados por la herramienta *Pajek*.

3.3.1. Gephi

Gephi es una herramienta de software open source para graficar y analizar redes (Bastian et al, 2009), está disponible para instalarse en plataforma *Windows*, *Linux*, y *Mac OS X*. Esta herramienta de visualización y análisis puede ser utilizada en redes sociales, *Web* semántica, seguridad nacional, mapeo de la *Web*, monitoreo de infraestructura, estudio de redes de negocios, investigación de interacción de genes (Gephi Consortium, 2010).

Para llevar a cabo el análisis y visualización de las redes *Gephi* tiene implementados varios algoritmos, entre ellos se encuentran el *ForceAtlas*, *Fruchterman-Reingold*, *Yifan Hu Multilevel*, *OpenOrd*, *ForceAtlas 2*, *Circular*, *Radial Axis*, y *Geographic Map*. Estos algoritmos están divididos en 4 categorías de acuerdo con la tipología que se quiera enfatizar en las redes, las categorías presentadas son: i) División, ii) Complemento, iii) *Ranking*, y iv) Representaciones Geográficas. En el Tabla 3.4 se muestra la clasificación y los algoritmos ligados a cada una de ellas (Gephi, 2011).

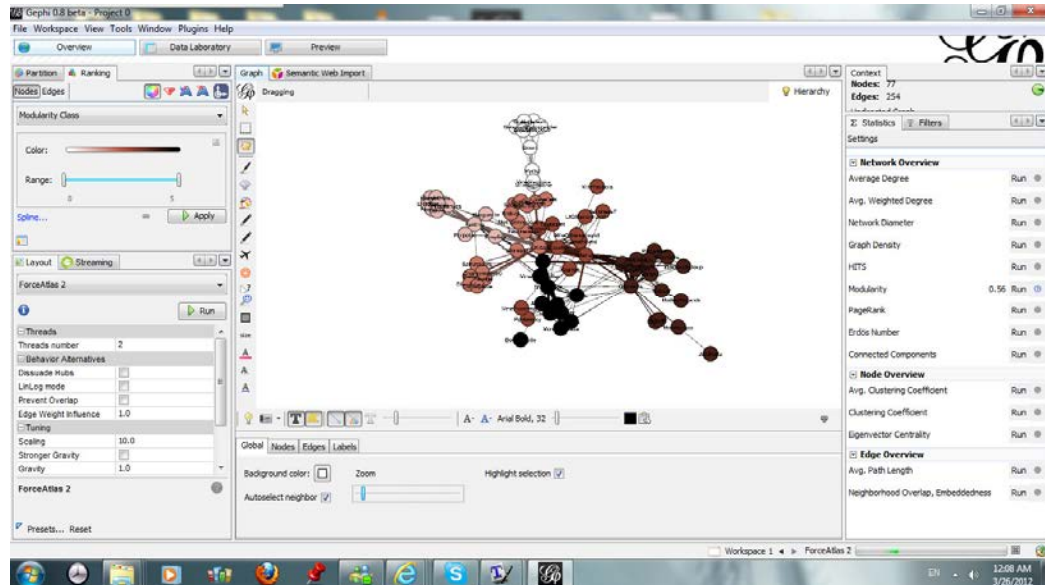


Figura 3.6: Herramienta de visualización *Gephi*.

Tabla 3.4: *Gephi*: Categorías y Algoritmos.

Categoría	Algoritmo
División	<i>OpenOrd</i>
Complemento	<i>ForceAtlas</i> , <i>ForceAtlas 2</i> , <i>Yifan Hu</i> , <i>Fruchterman- Reingold</i>
Ranking	<i>Circular</i> , <i>Radial Axis</i>
Representaciones geográficas	<i>GeoLayout</i>

El tipo de distribución obtenida con el algoritmo *ForceAtlas* de *Gephi*, es utilizado para espacializar redes *Small-World/Scale-free*. Este algoritmo se centra en la calidad, lo cual permite una interpretación rigurosa de la gráfica con el menor número de posibles sesgos, y una buena legibilidad; el algoritmo es del tipo fuerza-dirigido (Gephi, 2011), el algoritmo *ForceAtlas 2* es una versión mejorada del algoritmo *ForceAtlas* para trabajar con redes grandes (Jacomy et al, 2011).

Fuerza-dirigidos con algoritmos multinivel para reducir la complejidad de las redes. Este algoritmo se puede utilizar en redes de gran tamaño debido a que es muy rápido (Hu, 2005).

OpenOrd es un algoritmo basado en el algoritmo *Fruchterman-Reingold*, trabaja con un número controlado de iteraciones a través de la aplicación de un algoritmo de recocido simulado (este tipo de algoritmo está inspirado en la técnica del recocido de acero y cerámicas que consiste en calentar y después enfriar el material para variar sus propiedades físicas) (Martin et al, 2011).

El algoritmo *Circular* muestra los nodos de la gráfica ordenados en forma circular por identificación del nodo, grado, centralidad, o por atributo. Se utiliza para mostrar la distribución de los nodos y los enlaces entre ellos (Groeninger, 2010).

Radial Axis es un algoritmo que agrupa los nodos y los muestra en ejes que irradian hacia el exterior de un círculo central, los grupos son generados usando métricas de grado, centralidad o algún atributo (Groeninger, 2010). Con este algoritmo es posible analizar la estructura de una comunidad dentro de su red, permite visualizar homofilia de los nodos por grado, la distribución de los nodos por grado dentro de cada comunidad, se visualizan los enlaces dentro de las comunidades, así como los enlaces entre las comunidades (Gephi, 2011).

Los algoritmos de tipo geográfico son utilizados para mostrar la posición de los nodos de acuerdo con sus coordenadas geográficas (latitud/longitud). Existen diversos algoritmos de este tipo, entre ellos está el Proyecto Mercator (utilizado en *Google maps*) que es un algoritmo para elaborar planos terrestres.

3.3.2. Pajek

La herramienta de visualización *Pajek*, desarrollada por *Vladimir Batagelj* y *Andrej Mrvar*, es una herramienta ampliamente utilizada en el campo del análisis de redes sociales y está disponible para plataforma *Windows* y *Linux* vía *wine*. *Pajek* fue especialmente diseñado para soportar la abstracción mediante factorialización de

grandes redes en otras de menor tamaño, para proporcionar herramientas potentes para visualizar estas redes, además de implementar una selección de algoritmos para el análisis de grandes redes (Nooy, 2011). En su módulo de visualización *Pajek* (ver Figura 3.7) implementó los algoritmos *Circular*, *Fruchterman-Reingold*, y el *Kamada-Kawai*; este último algoritmo, al igual que el algoritmo *Fruchterman-Reingold*, es fuerza-dirigido y sigue el mismo principio del sistema de partículas.

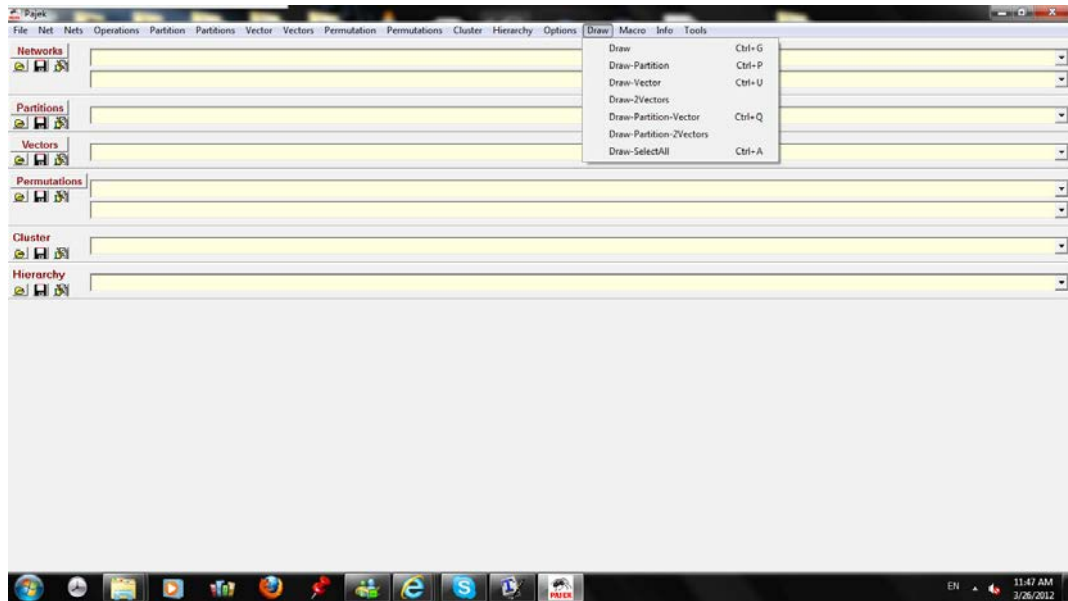


Figura 3.7: Herramienta de visualización *Pajek*

3.3.3. *VOSviewer*

VOSviewer es una aplicación que fue desarrollada principalmente para ser utilizada para el análisis de redes bibliométricas (Figura 3.8). Con *VOSviewer* se pueden crear mapas de las publicaciones, autores o revistas tomando como base redes de co-citación o bibliográficas, para crear mapas de palabras clave basadas en redes de co-ocurrencia. No obstante, el principal objetivo de esta herramienta es el análisis de redes bibliométricas, se puede utilizar para crear mapas basados en cualquier tipo de red. Además de que *VOSviewer* es usado para crear mapas basado en redes de datos, también permite visualizarlos y explorarlos. Con su módulo de visualización es posible ver diferentes aspectos de los mapas, cuenta con funcionalidades que permiten hacer

CAPÍTULO 4

DISEÑO E IMPLEMENTACIÓN DE UN MODELO TEÓRICO- METODOLÓGICO PARA LA INTEGRACIÓN DE DATOS BIBLIOGRÁFICOS DE PATENTES Y LA OBTENCIÓN DE SUS INDICADORES

4. Diseño de implementación de un modelo teórico-metodológico para la integración de datos bibliográficos de patentes y la obtención de sus indicadores

Este capítulo está dividido en tres secciones de resultados: en la primera se modela el *Data Warehouse* propuesto que integra las bases de datos *Derwent Innovation Index*, *Interpat*, y *Worldwide*; en la segunda se exponen los requerimientos funcionales que fueron tomados en cuenta para implementar los scripts que permiten transformar los datos, poblar la base de datos, y obtener reportes que pueden ser utilizados en las diferentes herramientas para realizar el análisis de los datos; en la tercera se muestra la aplicación de estos resultados en un conjunto de patentes españolas concedidas en el periodo 1997-2008.

4.1. Modelo teórico-metodológico de integración de datos bibliográficos de patentes

Para diseñar el Modelo de Integración de Bases de Datos Bibliográficos de Patentes o *Data Warehouse* se siguieron los pasos sugeridos por la Metodología de Descubrimiento de Conocimiento en Bases de Datos o proceso KDD (por sus siglas en inglés, *Knowledge Discovery in Databases*). En esta sección se muestra el modelo propuesto, así como la definición de cada una de las relaciones y atributos que lo conforman.

4.1.1. Selección de datos

Las tres bases de datos seleccionadas para construir el modelo de *Data Warehouse* son *Interpat* de INVENES, *Derwent Innovation Index*, y *Worldwide* de *espacenet*. Como se mencionó en el capítulo anterior, la base de datos *Interpat* es utilizada debido a la facilidad que presenta para obtener las patentes de un país a través de una base de datos nacional, además proporciona información más específica sobre los solicitantes de las patentes. De esta base de datos se seleccionó el título, tipo de documento, fecha de solicitud, fecha de publicación, fecha del informe del estado de

la técnica, fecha de concesión, fecha de publicación de la concesión, código de publicación, solicitantes, nacionalidad del solicitante, provincia del solicitante, dirección del solicitante, inventores, nacionalidad del inventor, número de publicación (número de la patente) y resumen. Los datos considerados de la base *Derwent Innovation Index* son el identificador único de la familia de patentes, los números de patentes, el título, nombre de los inventores, nombre de los solicitantes, patentes referenciadas.

Aun cuando la base de datos *Derwent Innovation Index* recoge información de las oficinas de patentes más importantes del mundo, no es posible determinar la nacionalidad del solicitante de la patente a través de ésta. Al analizar los datos proporcionados por la base de datos *Worldwide-espacenet* se encontró que en ésta sí está presente la nacionalidad de los solicitantes de las patentes, por esta razón se tomó la decisión de utilizar *Worldwide-espacenet* para trabajar con datos relacionados a las patentes citadas por los inventores y examinadores de las patentes de la base de datos *Derwent Innovation Index*. Los atributos que fueron tomados en cuenta de la base *Worldwide-espacenet* son: el número de publicación, fecha de publicación, solicitantes, y la clasificación internacional de patentes.

4.1.2 Modelo de Integración de bases de datos

Una vez seleccionadas las fuentes de datos, y analizados los atributos se procede a la integración de éstos en un solo repositorio para permitirle a las instituciones o corporaciones que harán uso de ellos tener una visión amplia de los datos, es decir, permitir que los analistas de datos puedan verlos como un colectivo de datos que proviene de una sola fuente bien definida y granulada (Inmon *et al*, 2008).

A partir de las bases de datos seleccionadas y con fundamento en la metodología del KDD se propone el modelo mostrado en la Figura 4.1. Como se puede observar, en el modelo se tienen tres fuentes de datos, donde la base de datos bibliográficos de patentes nacional sienta las bases a partir de las cuales se desprenden los procesos que se deben seguir en el modelo propuesto (1). A partir de estas tres fuentes de datos se

diseñó el modelo del repositorio de datos donde se almacenan los datos (2), mediante *scripts* CGI se procesan los datos obtenidos de la base de datos de patentes nacional (los cuales están almacenados en archivos de tipo texto), éstos son almacenados en el

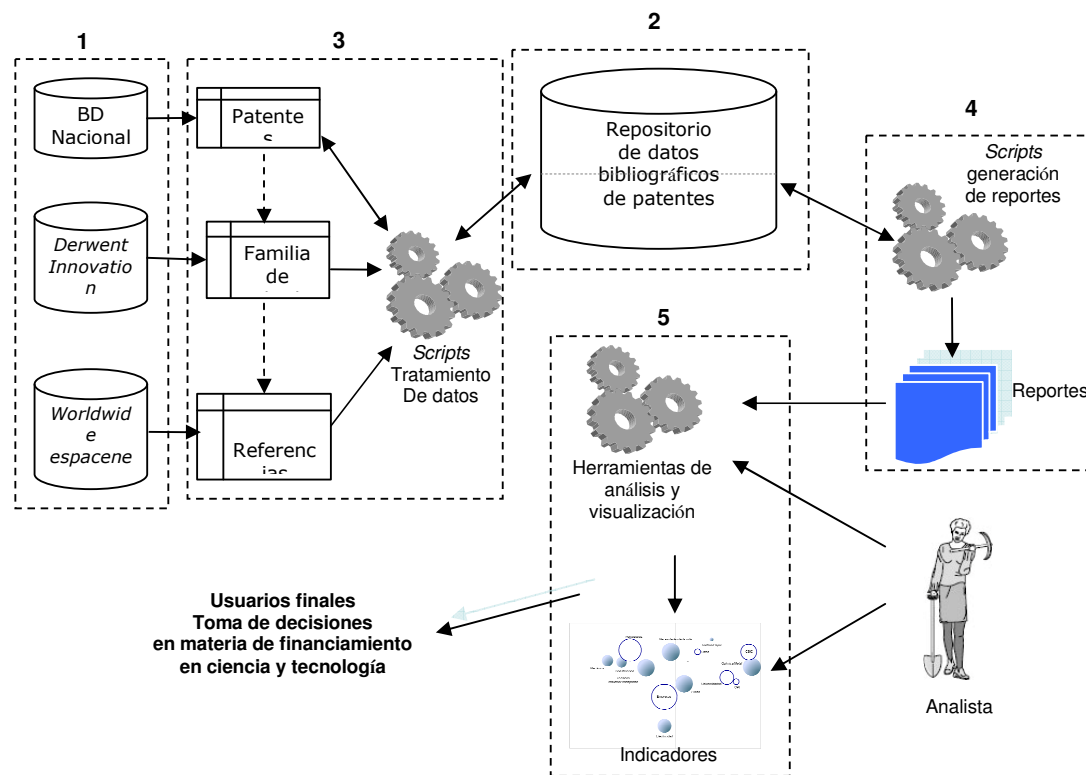


Figura 4. 1: Modelo de integración de datos bibliográficos de patentes.

repositorios de datos, posteriormente a partir de ellos y también mediante la utilización de *scripts* se obtienen los números de patentes que son descargados de la base de datos *Derwent Innovation Index* y almacenados en archivos. Los archivos que contienen los registros de las familias de patentes descargados de la base de datos *Derwent Innovation Index* son ingresados al repositorio de datos utilizando *script* desarrollados para ese propósito. Para finalizar con el proceso de carga de datos, se obtienen los números de patentes referenciadas por las familias de patentes, los cuales son extraídos de la base de datos *Worldwide-espacenet*, al igual que los registros de las anteriores bases de datos éstos son almacenados en archivos de tipo texto los cuales son insertados en el repositorio de datos. Al terminar con el proceso de carga de datos se realizan una serie de procesos para poner a punto de explotación los datos contenidos en el repositorio de datos bibliográficos de patentes (3).

Otros procesos de gran importancia en la metodología del KDD son los procesos relacionados con obtención de reportes para llevar a cabo la explotación de los datos. En este caso en particular los reportes son la fuente de datos que son utilizadas para la obtención de indicadores de patentes (4). En el modelo presentado en la Figura 4.1 se observa que se proporciona un grupo de *scripts CGI*s para la obtención de reportes, los cuales son los datos de entrada en las diferentes herramientas de análisis tales como las hojas de cálculo, *Xlstat*, *Pajek*, *Gephi*, *VOSviewer* mencionar algunas (5).

En el capítulo 3 se mencionó que el modelo de *Data Warehouse* o repositorio de datos bibliográficos de patentes se diseñó utilizando la metodología para el diseño de bases de datos relacionales. Al diseñar el modelo se siguió esta metodología, la cual indica que se debe hacer un diseño de base de datos normalizado para evitar la redundancia de datos e inconsistencia de información, pero también se tomó en cuenta que en los *Data Warehouse* no es posible seguir estrictamente las formas normales de la metodología relacional debido a que es necesario tener datos repetidos para llevarla a cabo (Inmon *et al*, 2008). Por lo anterior, en el diseño del *Data Warehouse* que se muestra en la Figura 4.2 se puede apreciar que en las diferentes relaciones que forman parte de este modelo hay atributos que contienen los mismos datos.

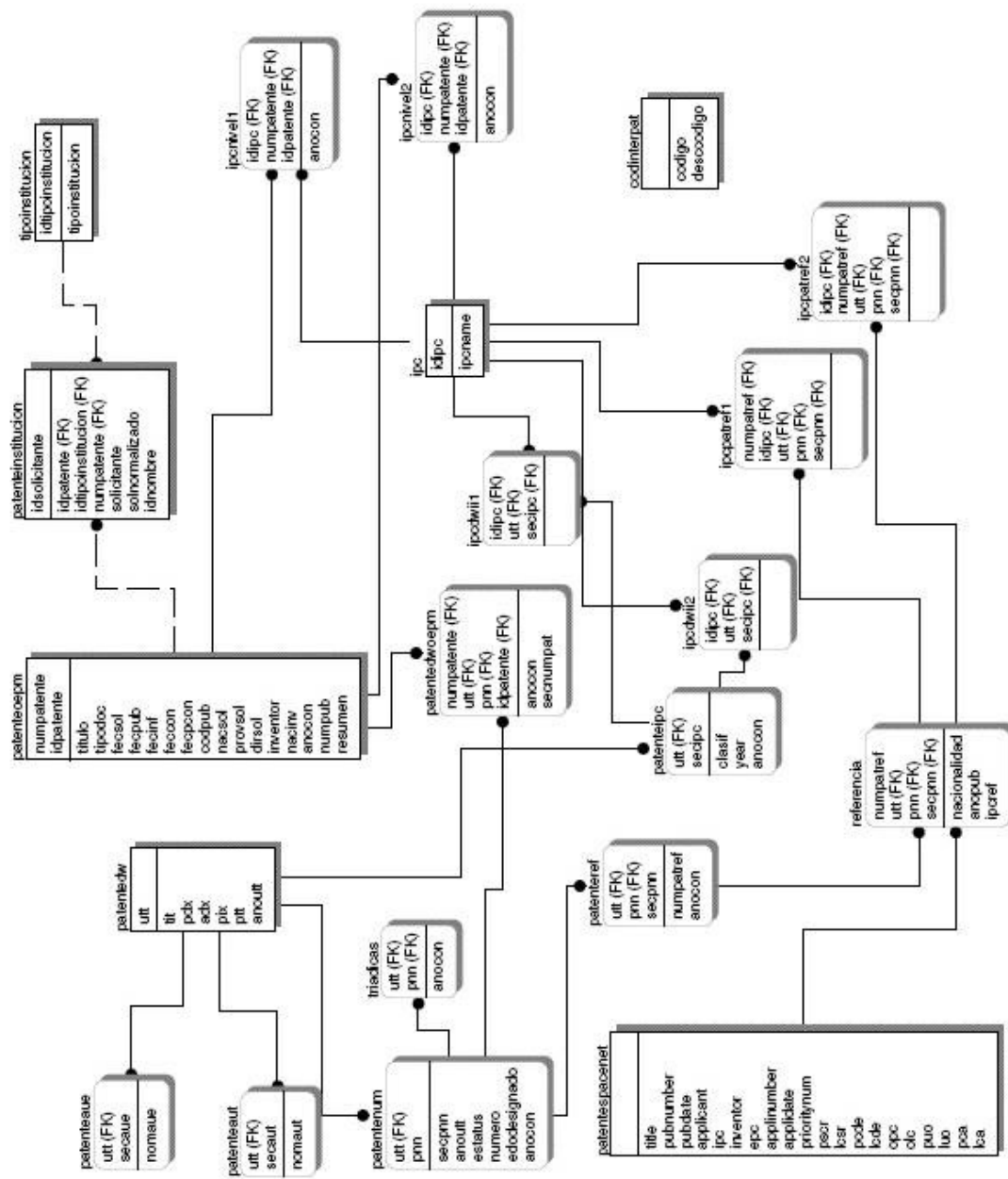


Figura 4.2: Diseño de la base de datos

4.1.3 Definición de relaciones y atributos

Un proceso de gran importancia en el modelado de base de datos es la definición de las relaciones y atributos integrantes del modelo para evitar la interpretación incorrecta o posibles ambigüedades, en la Tabla 4.1 se muestra la definición de las relaciones y atributos definidos en el modelo de integración diseñado. En este caso se indica el tipo de dato genérico de los atributos, es decir, tipo de datos numérico, alfabético o alfanumérico, también se indican los atributos que forman parte de la llave primaria de las relaciones escribiendo entre paréntesis las siglas PK (por las siglas en inglés de *Primary Key*), ampliamente reconocidas en la jerga informática.

Tabla 4.1: Definición de relaciones y atributos.

Relación/Atributos	Descripción
<i>Patenteoepm</i>	En esta relación se almacenan los registros de las patentes concedidas buscadas en la base de datos <i>Interpat</i> . Los números de patente de estos registros sirven de base para realizar la búsqueda en la base de datos <i>Derwent Innovation Index</i> .
<i>idpatente</i>	Atributo que identifica de forma única a cada registro (llave primaria). Es de tipo numérico secuencial (PK).
<i>numpatente</i>	Este atributo corresponde al número de patente concedido por la Oficina Española de Patentes y Marcas. Este número es generado siguiendo las normas establecidas por la OMPI, y es mediante éste que se puede identificar la patentes en las diferentes bases de datos mundiales. El atributo <i>numpatente</i> es de tipo alfanumérico de 8 caracteres. (PK)
<i>titulo</i>	El atributo <i>titulo</i> se corresponde con el título asignado por el o los solicitantes de la patente. Este atributo es de tipo texto.

<i>tipodoc</i>	Debido a que en la base de datos <i>Interpat</i> se almacenan tanto patentes como otro tipo de documentos denominados modelos de utilidad, en este atributo se indica que el registro es de tipo patente. El atributo es de tipo alfabético de un carácter, y el valor contenido es la letra P.
<i>fecsol</i>	Fecha de solicitud de la patente nacional (en OEPM). El atributo <i>fecsol</i> es de tipo alfanumérico de 8 caracteres, los primeros cuatro caracteres corresponden al año, los dos siguientes al mes, y los dos últimos indican el día.
<i>fecpub</i>	Fecha de publicación de la solicitud nacional (en OEPM). El tipo y formato de este atributo son iguales al atributo <i>fecsol</i> .
<i>feccon</i>	Fecha de concesión de la solicitud de patente nacional. El atributo <i>feccon</i> también es igual a los descrito para el atributo <i>fecsol</i> .
<i>fecpcon</i>	Fecha de publicación en BOPI de la concesión nacional (en OEPM). Este atributo también es de tipo alfanumérico de 8 caracteres, con los primeros cuatro caracteres indicando el año, dos para el mes, y los últimos el día.
<i>fecinftec</i>	Fecha de publicación del informe sobre el Estado de la Técnica. El atributo <i>fecinftec</i> también se define como los anteriores atributos relacionados con las diferentes fechas presentes en un documento de patente.
<i>codpub</i>	Código de publicación de la patente española, este código indica la situación jurídica de la patente. En el anexo 4 se proporciona más información sobre este código. El atributo es de tipo alfabético de 2 caracteres.

<i>nomsol</i>	Atributo utilizados para guardar copia del nombre del o los solicitantes de la patente, <i>nomsol</i> es de tipo alfabético.
<i>nacsol</i>	Este atributo es utilizado para almacenar una copia de la nacionalidad del solicitante de la patente. El atributo <i>nacsol</i> es de tipo alfabético, y se utiliza el código de dos letras de países definidos en la Norma ST.3 (ver anexo 1).
<i>provsol</i>	En este atributo se almacena copia de la provincia del solicitante. <i>provsol</i> es de tipo numérico.
<i>dirsol</i>	El atributo <i>dirsol</i> es utilizado para almacenar una copia de la dirección del solicitante, y es de tipo alfabético.
<i>inventor</i>	Este atributo tiene las mismas características del atributo <i>nomsol</i> , a diferencia que en éste se almacena copia de los nombres de los inventores del invento.
<i>nacinv</i>	Atributo definido de forma idéntica al atributo <i>nacsol</i> , con información relativa a la nacionalidad del inventor.
<i>anocon</i>	En el atributo <i>anocon</i> se almacena el año de concesión de la patente, es de tipo numérico de 4 posiciones, este dato se obtiene a partir de los valores del atributo <i>feccon</i> definido anteriormente.
<i>resumen</i>	En este atributo es utilizado para almacenar el resumen del documento de patente, éste es de tipo alfabético.
<i>tipoinstitucion</i>	En esta relación se almacena el catálogo de autoridad utilizado para clasificar a los solicitantes de las patentes por sector de actividad. Hasta el momento son 7 los diferentes tipos de sectores almacenados en esta relación (ver Figura 4.1).

<i>idtipoinstitucion</i>	El atributo <i>idtipoinstitucion</i> identifica de manera única a los registros almacenados en la relación, es la llave primaria, y es de tipo numérico (PK)
<i>tipoinstitucion</i>	En este atributo se almacena el nombre de cada uno de los diferentes sectores que conforman el catálogo de autoridad utilizado para clasificar a los solicitantes de patentes.
<i>patenteinstitucion</i>	En la relación <i>patenteinstitucion</i> se almacenan los datos relacionados con información de los solicitantes de las patentes.
<i>idsolicitante</i>	El atributo <i>idsolicitante</i> identifica de manera única a los registros almacenados en la relación, es la llave primaria, y es de tipo numérico (PK)
<i>idtipoinstitucion</i>	Este atributo es una llave foránea, ver definición en la relación <i>tipoinstitucion</i> .
<i>numpatente</i>	Este atributo es una llave foránea, ver definición en la relación <i>patenteoepm</i> .
<i>idpatente</i>	Este atributo es una llave foránea, ver definición en la relación <i>patenteoepm</i> .
<i>solicitante</i>	Atributo utilizado para guardar copia del nombre del o los solicitantes de la patente, <i>solicitante</i> es de tipo alfabético.
<i>solnormalizado</i>	En este atributo se almacena el nombre normalizado del solicitante.
<i>idnombre</i>	Identificador del nombre normalizado del solicitante.
<i>ipc</i>	La relación <i>ipc</i> es utilizada para almacenar el catálogo de autoridad de la Clasificación Internacional de Patentes, la cual es usada para clasificar la temática de las patentes en las oficinas de patentes.

<i>idipc</i>	En el atributo <i>idipc</i> se utiliza para almacenar el código asignado a cada una de las temáticas presentes en la Clasificación Internacional de Patentes. Este atributo es la llave primaria de la relación <i>ipc</i> , es de tipo alfanumérico. (PK)
<i>nombreipc</i>	Atributo en cual se almacenan los nombres descriptivos de los códigos utilizados en la Clasificación Internacional de Patentes. <i>nombreipc</i> es un atributo de tipo alfabético.
<i>ipcnivel1</i>	En la relación <i>ipcnivel1</i> se almacenan las clasificaciones a nivel clase o sección de las patentes.
<i>numpatente</i>	Llave foránea, ver definición en la relación <i>patentoeopm</i> . (PK)
<i>idpatente</i>	Llave foránea, ver definición en la relación <i>patentoeopm</i> . (PK)
<i>idipc</i>	Llave foránea, ver definición en la relación <i>ipc</i> . (PK)
<i>anocon</i>	Año de concesión de la patente, los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patentoeopm</i> . Ver definición de este atributo en la relación antes mencionada.
<i>ipcnivel2</i>	En la relación <i>ipcnivel2</i> se almacenan la clasificación a nivel subgrupo de las patentes.
<i>numpatente</i>	Llave foránea, ver definición en la relación <i>patentoeopm</i> . (PK)
<i>idpatente</i>	Llave foránea, ver definición en la relación <i>patentoeopm</i> . (PK)
<i>idipc</i>	Llave foránea, ver definición en la relación <i>ipc</i> . (PK)

<i>anocon</i>	Año de concesión de la patente. Los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i> . Ver definición de este atributo en la relación antes mencionada.
<i>patentedw</i>	En la relación <i>patentedw</i> se almacenan datos de las familias de patentes descargadas de la base de datos <i>Derwent Innovation Index</i> . Es importante hacer énfasis en que esta base de datos contiene familias de patentes, de tal forma que las patentes descargadas de la base de datos <i>Interpat</i> están contenidas dentro de estas familias, esto implica que dentro de una familia es posible encontrar una o más de las patentes bajadas de <i>Interpat</i> .
<i>utt</i>	Este atributo es la llave primaria de la relación. Los valores de éste son los asignados por <i>Derwent Innovation Index</i> , esta definido de tipo alfabético de 10 caracteres. (PK)
<i>tit</i>	El atributo <i>tit</i> es utilizado para almacenar el título del invento en inglés. Es de tipo alfabético.
<i>pdx</i>	Familia de patentes. Aquí se encuentran datos de cada uno de los miembros de la familia de patentes del registro, los datos mostrados son: el número de la publicación, la fecha de publicación, la clasificación internacional de patentes principal, la semana en la que el registro fue ingresado a la base de datos, número de páginas, y el idioma original de la patente. Este atributo es de tipo alfabético, y como se puede observar los datos aquí almacenados siguen conservando el formato original de la base de datos <i>Derwent Innovation Index</i> .

	<p><i>adx</i> Detalles de solicitud y fecha. En este atributo encontramos el número de patente asignado por la oficina de patentes local y la fecha. La fecha de solicitud es la fecha en la cual se hizo la solicitud en la oficina local. El atributo es de tipo alfabético, y el contenido conserva el formato de la fuente original.</p> <p><i>pix</i> En este atributo se almacenan datos relacionados con la solicitud prioritaria. El atributo <i>pix</i> es de tipo alfabético.</p> <p><i>ptt</i> Los datos almacenados en este atributo hacen referencia al tipo de documento. El atributo es de tipo alfabético de un carácter, y el valor contenido es el carácter P.</p> <p><i>anocon</i> Año de concesión de la patente, los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i>. Ver definición de este atributo en la relación antes mencionada.</p>
<i>patentenum</i>	<p>En esta relación se almacena datos relacionados con cada una de las patentes miembro de las familias de patentes descargadas de la base de datos <i>Derwent Innovation Index</i>. Como se mencionó anteriormente, los registros que proporciona la base de datos <i>Derwent</i> son patentes agrupadas en familias.</p>
	<p><i>utt</i> Llave foránea, ver definición en la relación <i>patentedw</i>. (PK)</p> <p><i>pnn</i> Este atributo se utiliza para almacenar el número de patente que forma parte de la familia. El formato de los datos almacenados aquí está compuesto por cuatro elementos que son: 1) el código de 2 letras</p>

	<p>correspondiente al estado designado en donde se registró la patente, 2) un conjunto de números, 3) el símbolo -, y 4) el código que indica el estatus jurídico de la patente. Es definido de tipo alfanumérico. (PK)</p>
<i>secpnn</i>	<p>Este es un atributo que indica la posición en la que se encuentra colocado este miembro dentro de la familia de patentes proporcionadas por <i>Derwent Innovation Index</i>. El atributo es de tipo numérico.</p>
<i>anocon</i>	<p>Año de concesión de la patente, los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i>. Ver definición de este atributo en la relación antes mencionada.</p>
<i>edodesignado</i>	<p>En este atributo se almacena el código del estado designado en el que fue solicitada la patente. Es el mismo estatus que se encuentra registrado como parte del número de patente en el atributo <i>pnn</i>. El atributo <i>edodesignado</i> es de tipo alfabético.</p>
<i>numero</i>	<p>El atributo número es utilizado para almacenar el segundo elemento de un número de patente, <i>número</i> es de tipo alfabético.</p>
<i>estatus</i>	<p>En este atributo se almacena el código que indica el estatus jurídico de la patente. Es el mismo estatus que se encuentra registrado como parte del número de patente en el atributo <i>pnn</i>, <i>estatus</i> es de tipo alfabético.</p> <p>Como se puede observar, los tres atributos anteriores son la descomposición de los números de patente en cada una de sus partes.</p>
<i>Patentedwoepm</i>	<p>En es la relación mediante la cual se hace el enlace de las</p>

	patentes extraídas de la bases de datos local Interpat y las familias de patentes obtenidas de la base de datos <i>Derwent Innovation Index</i> .
<i>numpatente</i>	Llave foránea, ver definición en la relación <i>patenteoepm</i> . (PK)
<i>idpatente</i>	Llave foránea, ver definición en la relación <i>patenteoepm</i> . (PK)
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>pnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> . (PK)
<i>anocon</i>	Año de concesión de la patente. Los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i> . Ver definición de este atributo en la relación antes mencionada.
<i>secpnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> .
<i>patenteaue</i>	En esta relación se almacenan datos relacionados con los solicitantes de las patentes, como son el nombre del solicitante y el orden en que se encuentran en la lista proporcionada por <i>Derwent Innovation Index</i> .
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>secaue</i>	Este es un atributo que indica la posición en la que se encuentra colocado el nombre del solicitante en la lista de solicitantes proporcionada por <i>Derwent Innovation Index</i> . El atributo es de tipo numérico. (PK)

<i>nomaue</i>	En el atributo <i>nomaue</i> se almacena el nombre del solicitante de la patente, el atributo es de tipo alfabético.
<i>patenteaut</i>	En esta relación se almacenan datos relacionados con los inventores de las patentes, como son el nombre del solicitante y el orden en que se encuentran en la lista proporcionada por <i>Derwent Innovation Index</i> .
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>secaut</i>	Este es un atributo que indica la posición en la que se encuentra colocado el nombre del inventor en la lista de solicitantes proporcionada por <i>Derwent Innovation Index</i> . El atributo es de tipo numérico. (PK)
<i>nomaut</i>	En el atributo <i>nomaut</i> se almacena el nombre del inventor de la patente. El atributo es de tipo alfabético.
<i>patenteipc</i>	La relación <i>patenteipc</i> es utilizada para almacenar datos relacionados con las temáticas correspondientes a la familia de patentes, un ejemplo del formato de la clasificación temática es el siguiente: B60H-001/18.
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>secipc</i>	Este es un atributo que indica la posición en la que se encuentra colocada la temática de la patente dentro de la lista de temáticas que le fueron asignadas a la familia de patentes. Este ordenamiento es proporcionado por <i>Derwent Innovation Index</i> . El atributo es de tipo numérico. (PK)

<i>ipcfamilia</i>	El atributo <i>ipcfamilia</i> es utilizado para almacenar la clasificación temática de la familia de patentes. Este atributo es de tipo alfabético.
<i>anocon</i>	Año de concesión de la patente. Los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i> . Ver definición de este atributo en la relación antes mencionada.
<i>ipcdwii1</i>	En la relación <i>ipcdwii1</i> se almacenan las clasificaciones a nivel clase o sección de la familia de patentes.
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>secipc</i>	Llave foránea, ver definición en la relación <i>patenteipc</i> . (PK)
<i>idipc</i>	En este atributo se almacena el código a nivel clase de la familia de las patentes, y es el atributo mediante el cual se relaciona las relaciones <i>patenteipc</i> e <i>ipc</i> . Ver definición del atributo <i>idipc</i> en la relación <i>ipc</i> . (PK)
<i>ipcdwii2</i>	En la relación <i>ipcdwii2</i> se almacenan las clasificaciones a nivel subgrupo de la familia de patentes.
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>secipc</i>	Llave foránea, ver definición en la relación <i>patenteipc</i> . (PK)
<i>idipc</i>	En este atributo se almacena el código a nivel subgrupo de la familia de las patentes, y es el atributo mediante el cual se relaciona las relaciones <i>patenteipc</i> e <i>ipc</i> . Ver definición del atributo <i>idipc</i> en la relación <i>ipc</i> . (PK)

<i>patenteref</i>		En esta relación se almacenan datos relacionados con las patentes referenciadas por los inventores y/o examinadores. Es importante hacer énfasis en que la base de datos <i>Derwent Innovation Index</i> proporciona las patentes referenciadas, pero no es posible determinar las patentes referenciadas por los inventores de las patentes referenciadas por los examinadores.
	<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
	<i>pnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> . (PK)
	<i>secpnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> .
	<i>numpatref</i>	En este atributo se almacena el número de patente referenciado, el formato es igual al formato de los números de patentes citantes, ver definición del atributo <i>pnn</i> en la relación <i>patentenum</i> .
<i>triadica</i>		En esta relación se almacenan el identificador único de las familias de patentes triádicas, y el año de concesión. La selección de estas familias se hizo tomando como criterio que tuviera dentro de sus miembros números de patentes con código de estado designado US, EP, y JP que corresponden a la Oficina Americana de Patentes, a la Oficina Europea de Patentes, y a la Oficina Japonesa de Patentes respectivamente.
	<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
	<i>anocon</i>	Este atributo indica el año de concesión de la patente, los datos aquí contenidos son iguales a los del atributo <i>anocon</i> de la relación <i>patenteoepm</i> .

<i>patentespacenet</i>	La relación <i>patentespacenet</i> es donde se almacenan los datos extraídos de la base de datos <i>Worldwide-espacenet</i> , éstos corresponden a las patentes referenciadas por las familias de patentes. Los datos son utilizados para realizar un preprocesamiento para extraer los datos que son utilizados en el modelo y poblar las relaciones que almacenan los datos referentes a las patentes referenciadas. En el diseño de la base de datos se observan más atributos de los definidos aquí, esto se debe a que un gran número de ellos no tienen datos (como ya se mencionó en el capítulo 3), por lo tanto solo se definen los utilizados en el modelo aquí propuesto.
<i>pubnumber</i>	En el atributo <i>pubnumber</i> se almacena el número de patente referenciado. La definición de este atributo es igual a la definición del atributo <i>pnn</i> en la relación <i>patentenum</i> . Es mediante <i>pubnumber</i> que se interrelacionan las patentes citantes y las referencias. El formato de los datos es similar al formato de los datos almacenados en el atributo <i>pnn</i> , pero en este caso el código que indica el estatus jurídico de la patente se encuentra entre paréntesis.
<i>pubdate</i>	El atributo <i>pubdate</i> hace referencia a la fecha de publicación de la patente. El formato de esta fecha es año (4 dígitos), mes (2 dígitos), y día (2 dígitos) separados por un guión (-). El atributo es de tipo alfabético.
<i>applicant</i>	En este atributo se almacena el nombre de los solicitante de la patente, los nombres están separados por el carácter “;”, para cada solicitante se indica, entre paréntesis cuadrados, la nacionalidad, utilizando el código de dos caracteres de los países. Este atributo es de tipo alfabético.

<i>ipc</i>	En el atributo <i>ipc</i> se almacenan los códigos de las temáticas asignadas a la patente. Estos códigos están separados por el carácter “;”. El formato de los datos aquí almacenados es igual al formato de los datos almacenados en el atributo <i>ipcfamilia</i> de la relación <i>patenteipc</i> .
referencia	Esta relación es utilizada para almacenar los datos de las patentes referenciadas, éstos se obtienen del procesamiento que se realiza a los datos almacenados en la relación <i>patentespacenet</i> , y la relación <i>patenteref</i> . Los atributos que permiten identificar la interrelación entre estas dos relaciones son <i>pubnumber</i> y <i>numpatref</i> , <i>patentespacenet</i> y <i>patenteref</i> respectivamente.
<i>numpatref</i>	En este atributo se almacena el número de la patente referenciado. Este dato es el mismo que está almacenado en el atributo <i>numpatref</i> de la relación <i>patenteref</i> , y en el atributo <i>pubnumber</i> de la relación <i>patentespacenet</i> . El atributo es de tipo alfabético. (PK)
<i>utt</i>	Llave foránea, ver definición en la relación <i>patentedw</i> . (PK)
<i>pnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> . (PK)
<i>secpnn</i>	Llave foránea, ver definición en la relación <i>patentenum</i> . (PK)
<i>nacionalidad</i>	En este atributo se almacena la nacionalidad del solicitante, es el código de dos letras que se encuentra indicado entre paréntesis cuadrados en el atributo <i>applicant</i> de la relación <i>patentespacenet</i> . El atributo es de tipo alfabético.

<i>anopub</i>	En el atributo <i>anopub</i> se almacena el año de publicación de la patente referenciada. Este dato se obtiene del atributo <i>pubdate</i> de la relación <i>patentespacenet</i> . El atributo es de tipo numérico.
<i>ipcref</i>	El atributo <i>ipcref</i> es utilizado para almacenar la clasificación temática de la patente referenciada. Este atributo es de tipo alfabético.
<i>ipcpatref1</i>	En la relación <i>ipcpatref1</i> se almacenan las clasificaciones a nivel clase de las patentes referenciadas.
<i>numpatref</i>	Llave foránea, ver definición en la relación <i>referencia</i> . (PK)
<i>utt</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)
<i>pnn</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)
<i>secpnn</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)
<i>idipc</i>	En este atributo se almacena el código a nivel clase de la patente referenciada, está definido de tipo alfabético.
<i>ipcpatref2</i>	En la relación <i>ipcpatref2</i> se almacenan las clasificaciones a nivel subgrupo de las patentes referenciadas.
<i>numpatref</i>	Llave foránea, ver definición en la relación <i>referencia</i> . (PK)
<i>utt</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)

<i>pnn</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)
<i>secpnn</i>	Llave foránea, ver definición en la relación <i>patenteref</i> . (PK)
<i>idipc</i>	En este atributo se almacena el código a nivel subgrupo de la patente referenciada, está definido de tipo alfabético.

Como se puede observar, hasta el momento se ha hecho selección de tres fuentes de datos, una vez realizada esta selección se determinaron los datos y por consiguiente los atributos que han sido utilizados para modelar un *Data Warehouse* de datos bibliográficos de patentes, estas actividades o procesos se corresponden con los pasos mencionados en la metodología del KDD. En la siguiente sección se describe el método usado para realizar el tratamiento de los datos extraídos de las diferentes bases de datos seleccionadas para almacenarlos en el repositorio, así como para generar reportes que son utilizados para la obtención de indicadores sobre patentes.

4.2. Tratamiento de datos

Los datos se trataron utilizando *scripts CGI*, tal como se menciona en la metodología, dividiéndolos en tres grupos a partir de sus requerimientos funcionales, los cuales son definidos en esta sección. El primer grupo de *scripts* está relacionado con la carga de los datos en la base de datos que es utilizada como repositorio; el segundo grupo son los *scripts* relacionados con la normalización, y el tercero agrupa a los *scripts* relacionados con la generación de reportes que permitirán a los responsables del análisis, utilizar los datos en las diferentes herramientas de análisis y visualización. Para cada uno de estos tres grupos se planteo un objetivo.

4.2.1. Carga de los datos

El propósito de desarrollar un módulo que pueda ser utilizado como herramienta de apoyo para almacenar los datos extraídos de las bases de datos *Interpat*, *Derwent Innovation Index*, y *Worldwide-espacenet*, en la base de datos diseñada en el modelo de integración de bases de datos de patentes.

Tomando en consideración el objetivo descrito, a continuación se enumeran las características funcionales que debe tener este módulo:

- i. El proceso de carga de datos se inicia con los datos extraídos de la base de datos de la oficina local. En el capítulo anterior se mencionó que la base de datos local es *Interpat*, y que la interfaz permite descargar los registros buscados en archivos en el formato del procesador de textos *Word*.

Lo anterior conlleva a describir un proceso que el usuario de este módulo deberá realizar cuando descargue los datos de la base de datos *Interpat*. Este proceso consiste en convertir los archivos en formato *Word* a formato “texto sin formato”, para lo cual se debe abrir el archivo con el procesador de palabras *Word*. Una vez abierto se debe seleccionar la opción “guardar como” y guardar el archivo como tipo “texto sin formato”. Además de estas opciones,

el procesador de palabras *Word* muestra una ventana en donde se escoge el tipo de codificación que será utilizada para la conversión del archivo en formato *Word* a formato “Texto sin formato”, de las opciones mostradas en esta ventana se debe seleccionar la opción “*Windows* (predeterminada)”, seguida de la opción aceptar para guardar el archivo. Este nuevo formato quedará guardado en un archivo con extensión “txt”, en las Tablas 4.2 y 4.3 se muestra un fragmento de archivos con registros de la base de datos *Interpat* (descargados en diferentes años), convertido a formato texto. Este tipo de archivos será el que se utilizará en el proceso de carga de datos.

- ii. Crear el catálogo de autoridad con los nombres de los atributos presentes en los archivos extraídos de la base de datos *Interpat*. Los nombres de los atributos de los datos son extensos y tras haberlos analizado se observó que esas etiquetas pueden variar en el tiempo (ver Tablas 4.2 y 4.3).
- iii. Crear un catálogo de autoridades de la Clasificación Internacional de Patentes.
- iv. Reetiquetar los datos con los códigos del catálogo de autoridades.
- v. Ingresar los datos correspondientes al número de publicación, título, provincia del solicitante, nombre del solicitante de la patente, y fecha de concesión de la patente.
- vi. Generar una lista de los números de patentes que fueron ingresados a la base de datos para que el usuario busque esos registros en la base de datos *Derwent Innovation Index*.
- vii. Ingresar a la base de datos los registros de patentes extraídos de la base de datos *Derwent Innovation Index*.
- viii. Generar una lista con los números de patentes citadas por las patentes extraídas de la base de datos *Derwent Innovation Index*.
- ix. Buscar y descargar en *Worldwide-espacenet* los números de patentes resultantes de la lista (patentes referenciadas).
- x. Ingresar a la base de datos los registros de patentes extraídos de la base de datos *Worldwide*.

Tabla 4.2: Fragmento de un archivo de tipo texto sin formato. Fecha de descarga 2008.

```

REFERENCIA      : P9100992
TITULO          : SISTEMA DE MULTICONFERENCIA.
TIPO PATENTE    : P
FEC SOL OEPM    : 19910418
FEC PUB SOL OEPM : 19941101
FEC INF TEC     : 19941101
NUM SOL OEPM    : P9100992
FEC CON OEPM    : 19950330
FEC PUB CON OEPM : 19950501
NUM PUB OEPM    : 2059221
COD PUBLICACION : A1 B1
SOLICITANTES    : APLICACIONES ELECTRONICAS QUASAR, S.A.
DIRECCION SOLIC : REY PASTOR, 36 P.I. LEGANES,LEGANES 28914 MADRID
PROVINCIA SOL   : 28
NACION SOLICITANTE : ES
INVENTORES      : DE LA FUENTE DE LA FUENTE, ROGELIO
NACION INVENTOR : ES
CATEGORIA       : X,X,X,X,A,A
CLASIF A Y B    : H04M3/56
CLASIF 8        : H04M3/56 (2006.01)
INF EST TEC OEPM : GB2184324 GB2220546 FR2530903 GB2139055 WO8603912 US4045615
RESUMEN         : SISTEMA DE MULTICONFERENCIA. EL SISTEMA OBJETO DE LA INVENCION, CUYA
FINALIDAD ES LA REALIZACION DE VARIAS CONFERENCIAS SIMULTANEAS, MINIMIZANDO LA PRESENCIA
DE LA SEÑAL TRANSMITIDA EN EL CANAL DE RECEPCION CORRESPONDIENTE, ESTA CONSTITUIDO POR LA
ASOCIACION FUNCIONAL DE UNA SERIE DE HIBRIDAS DIGITALES (1) QUE SIRVEN DE LAZO DE UNION
ENTRE UNA MATRIZ DE CONMUTACION (2) CON UNIDADES DE CONTROL REMOTO (3). LA FUNCION DE
CONTROL LA REALIZA UNA UNIDAD MICROCONTROLADORA (4) A TRAVES DE UN BUS DE CONTROL (13) Y
DE UN INTERFACE (5) PARA HIBRIDAS (1) / MATRIZ DE CONMUTACION (2) Y UNIDADES DE CONTROL
REMOTO (3) RESPECTIVAMENTE.

-----
REFERENCIA      : P9100991
TITULO          : HIBRIDA DIGITAL PARA MULTICONFERENCIA.
TIPO PATENTE    : P
FEC SOL OEPM    : 19910418
FEC PUB SOL OEPM : 19930801
FEC INF TEC     : 19941101
NUM SOL OEPM    : P9100991
FEC CON OEPM    : 19950330
FEC PUB CON OEPM : 19950501
NUM PUB OEPM    : 2038887
COD PUBLICACION : A2 R B1
SOLICITANTES    : APLICACIONES ELECTRONICAS QUASAR, S.A.
DIRECCION SOLIC : REY PASTOR, 36 P.I.,LEGANES 28914 MADRID
PROVINCIA SOL   : 28
NACION SOLICITANTE : ES
INVENTORES      : DE LA FUENTE DE LA FUENTE, ROGELIO
NACION INVENTOR : ES
CATEGORIA       : Y,Y,A,A,A,A
CLASIF A Y B    : H04M3/56 H04B3/20
CLASIF 8        : H04B3/20 (2006.01) H04M3/56 (2006.01)
INF EST TEC OEPM : EP375015 EP122594 EP174749 EP192359 EP333581 GB2220546
RESUMEN         : HIBRIDA DIGITAL PARA MULTICONFERENCIA. LA HIBRIDA OBJETO DE LA INVENCION SE
TRATA DE UN SISTEMA DE CANCELACION DE ECO ELECTRICO POR FILTRADO DIGITAL ADAPTATIVO Y
PROCESADO DIGITAL DE LA SEÑAL, EL CUAL REALIZA UNA ADAPTACION AUTOMATICA, EN TIEMPO REAL,
A LOS PARAMETROS DE LINEA ENTRE TODO EL MARGEN, DE CIRCUITO ABIERTO A CORTOCIRCUITO,
ADAPTANDOSE INCLUSO A VARIACIONES INSTANTANEAS DE LA IMPEDANCIA DE LA LINEA. ESTA
CONSTITUIDA POR LA ASOCIACION FUNCIONAL DE UNA HIBRIDA ANALOGICA (1) Y UN FILTRO
ADAPTATIVO (13) CON LA COLABORACION DE CONVERTIDORES A/D (3, 11) Y D/A (7, 16), DE ADAPTACION.

```

Tabla 4.3: Fragmento de un archivo de tipo texto sin formato. Fecha de descarga 2010.

Referencia	: P200502938
Título	: COMPOSICION Y PROCEDIMIENTO PARA EXTRAER BETA-LACTOGLOBULINA DE MUESTRAS PROCESADAS Y NO PROCESADAS Y METODO DE INMUNOENSAYO PARA BETA-LACTOGLOBULINA.
Tipo de documento	: P
Fecha de solicitud nacional (en OEPM):	20051128
Fecha de publicación de la solicitud nacional (en OEPM):	20091014
Fecha de publicación del Informe sobre el Estado de la Técnica:	20091014
Número de solicitud nacional (en OEPM):	P200502938
Fecha de concesión de la solicitud nacional (en OEPM):	20100623
Fecha de publicación en BOPI de la concesión nacional (en OEPM):	20100708
Número de publicación de la solicitud nacional (en OEPM):	2326569
Código de publicación nacional:	A1 B1
Solicitantes	: CONSEJO SUPERIOR INVESTIG. CIENTIFICAS
Dirección del solicitante:	SERRANO 117,MADRID 28006
Código de la provincia del solicitante:	28
Nacionalidad del solicitante:	ES
Inventores	: DE FRUTOS GOMEZ,MERCEDES PELAEZ LORENZO,CRISTINA PUERTA GARCIA-BARROSO,ANGEL DE LA
Código de nacionalidad del inventor:	ES
Clasificación europea:	A23J3/08 G01N33/04 G01N33/551
Clasificación de Invención (8ª en adelante):	A23J3/08 (2006.01) G01N33/04 (2006.01)
Informe sobre el estado de la técnica de la OEPM (documentos citados en el IET):	* 1 REF.* * 1 REF.* * 1 REF.* * 1 REF.*
Categoría de los documentos nombrados en el IET:	X A,X A,X,Y,Y
RESUMEN	: Composición y procedimiento para extraer beta-lactoglobulina de muestras procesadas y no procesadas y método de inmunoensayo para beta-lactoglobulina.#Composición y procedimiento para extraer beta-lactoglobulina de muestras y un método de inmunoensayo para el análisis de beta-lactoglobulina ({be}-lactoglobulina, {be}-LG). Se refiere a una composición y un procedimiento para extraer la {be}-lactoglobulina ({be}-LG) existente en productos, que consiste en el empleo de una mezcla de un agente disruptor de puentes disulfuro y un agente desnaturalizante en un tampón de pH neutro o ligeramente básico, seguido por una etapa de tratamiento con una solución salina. Pueden ser empleados para extraer {be}-LG contenida en muestras de distintos orígenes sometidas o no a distintos procesos, como etapa previa a su análisis mediante inmunoensayo. Dicho método es válido para el análisis de productos, en particular para los destinados a personas alérgicas a {be}-LG. La invención se refiere también al análisis de {be}-LG y en especial al análisis mediante inmunoensayo y más concretamente mediante un ensayo en inmunoabsorbentes con detección enzimática (ELISA).

Para cumplir con los requerimientos funcionales antes descritos se deben realizar ciertas actividades comunes a ellos, debido a esta situación se creó un diagrama de secuencia general que las engloba. En la Figura 4.3 se observa que el usuario o analista de datos entra en la página inicial del sistema que permite realizar operaciones en el repositorio de datos. Esta primera página es una página *Web*

generada por un *script*, en ella se muestran dos opciones que corresponden a las actividades de tratamiento de datos y generación de reportes (ver Figura 4.4). En el diagrama de secuencia general (Figura 4.3) el analista de datos selecciona la opción de tratamiento de datos, esta tiene como objetivo hacer la invocación de un *script* que genera una página *Web* mostrándole una lista con los diferentes tipos de procesamiento que se deben realizar para el tratamiento de datos. Cada uno de estos procesos es realizado en el momento en que el analista los selecciona. En los epígrafes siguientes se describen los detalles específicos de cada uno de los *script* asociados a los procesos de la lista mostrada en la página de tratamiento de datos.

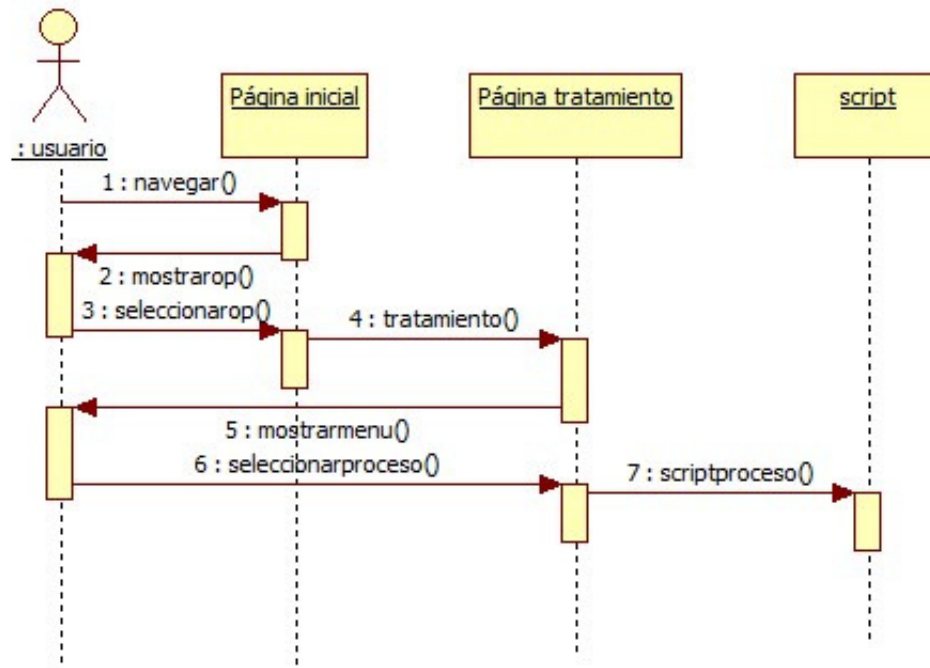


Figura 4.3: Diagrama general de tratamiento de datos.

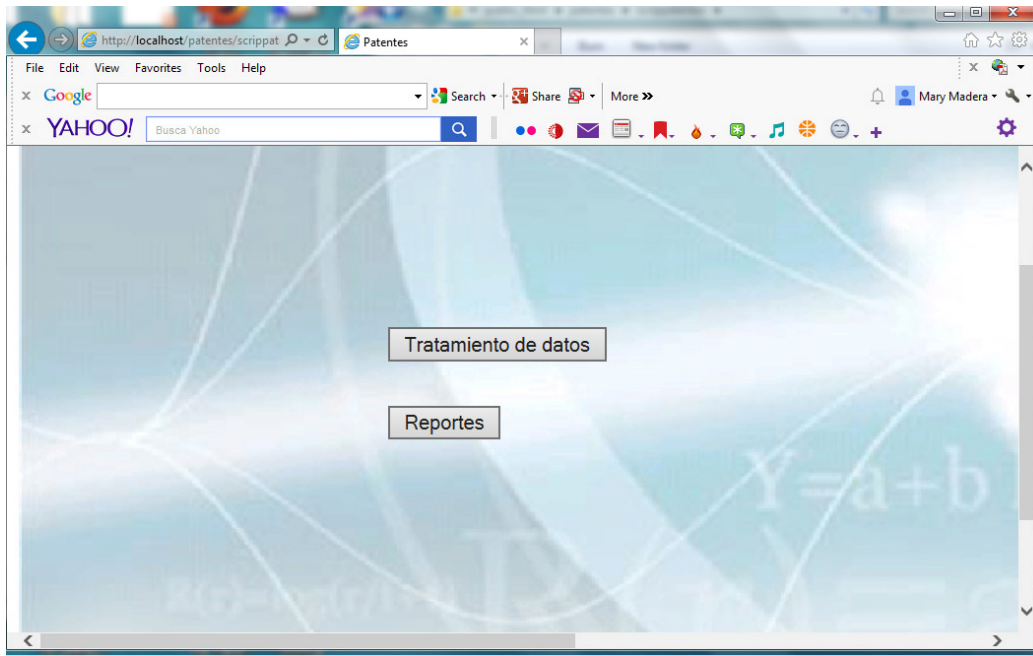


Figura 4.4: Página inicial.

4.2.1.1. Catálogos de autoridad

El catálogo de autoridad de los atributos de *Interpat* se genera con los nombre de los atributos de los datos de *Interpat* en formato texto, se proporciona con el esquema de la base de datos para que sea almacenado después de la creación de la base de datos en el administrador de bases de datos MySQL. En la Figura 4.5 se muestra un fragmento del contenido de este catálogo.

El catálogo de autoridades con los códigos y descripción de ellos se genera utilizando los datos de la Organización Mundial de la Propiedad Intelectual, accesibles a través de su sitio web¹. En la Figura 4.6 se muestra un fragmento de este catálogo.

¹ <http://cip.oepm.es/ipcpub/#lang=es&menulang=ES&refresh=page>

codigo ▲	desccodigo
CA	Categoría de los documentos nombrados en el IET
CE	Clasificación europea
CI	Clasificación adicional e indexada (8ª en adelan...
CL	Clasificación de invención anterior a octava edi...
CP	Código de publicación nacional
CS	Clasificación de Invención (8ª en adelante)
DS	Dirección del solicitante
ER	-----
FC	Fecha de concesión de la solicitud nacional (en O...
FE	Fecha de publicación en BOPI de la concesión nac...
FI	Fecha de publicación del Informe sobre el Estado ...
FP	Fecha de publicación de la solicitud nacional (en...
FS	Fecha de solicitud nacional (en OEPM)
IE	Informe sobre el estado de la técnica de la OEPM ...
IN	Inventores
NA	Nacionalidad del solicitante
NI	Código de nacionalidad del inventor
NP	Número de publicación de la solicitud nacional (...)
NS	Número de solicitud nacional (en OEPM)
PR	Prioridades
PS	Código de la provincia del solicitante
PT	Referencia
PT	Referencia
RE	RESUMEN
SO	Solicitantes

Figura 4.5: Catálogo de autoridad de atributos *Interpat*.

idipc	ipcname
A	NECESIDADES CORRIENTES DE LA VIDA
A01	AGRICULTURA; SILVICULTURA; CRIA; CAZA; CAPTURA; PE...
A21	COCCI?N EN HORNO; EQUIPAMIENTO PARA LA PREPARACI?N...
A22	CARNICERIA; TRATAMIENTO DE LA CARNE; TRATAMIENTO D...
A23	ALIMENTOS O PRODUCTOS ALIMENTICIOS; SU TRATAMIENTO...
A24	TABACO; PUROS; CIGARRILLOS; ARTICULOS PARA FUMADOR...
A41	VESTIMENTA
A42	SOMBRERERIA
A43	CALZADOS
A44	MERCERIA; JOYERIA
A45	OBJETOS DE USO PERSONAL O ARTICULOS DE VIAJE
A46	CEPILLERIA
A47	MOBILIARIO
A61	CIENCIAS M?DICAS O VETERINARIAS; HIGIENE
A62	SALVAMENTO; LUCHA CONTRA INCENDIOS
A63	DEPORTES; JUEGOS; DISTRACCIONES
A99	MATERIA NO PREVISTA EN OTRO LUGAR DE ESTA SECCI?N
B	TECNICAS INDUSTRIALES DIVERSAS; TRANSPORTES
B01	PROCEDIMIENTOS O APARATOS FISICOS O QUIMICOS EN GE...
B02	TRITURACION, REDUCCION A POLVO O DESINTEGRACION; T...
B03	SEPARACION DE SOLIDOS POR UTILIZACION DE LIQUIDOS ...
B04	APARATOS O MAQUINAS CENTRIFUGAS UTILIZADAS PARA LO...

Figura 4.6: Catálogo de autoridad de la Clasificación Internacional de Patentes.

4.2.1.2. Reetiquetado de datos

El reetiquetado de datos es necesario pues como se mencionó anteriormente, presentando cambios que pueden dificultar el flujo del proceso de ingreso a la base de datos (nombres largos y nombres que varían con el tiempo).

A continuación se describen los requerimientos funcionales para realizar el reetiquetado de los archivos en formato de texto de *Interpat*. Se cambian los nombres

largos de los atributos para facilitar el proceso de ingreso de los registros contenidos en estos archivos a la base de datos.

Precondiciones. Los archivos de *Interpat* en formato *Word* se deben convertir a formato texto mediante el proceso descrito al inicio de la sección de carga de datos.

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Solicitar nombre del archivo *Interpat* que será reetiquetado.
- v. Cambiar cada uno de los nombres de los atributos presentes en el archivo por su correspondiente código.
- vi. Generar un nuevo archivo en formato texto, contiene los mismos datos pero ahora con etiquetas del catálogo de autoridad.
- vii. Condición de error: si no se encuentra el archivo indicado, se envía mensaje de error.

En la Figura 4.7 se muestra la secuencia que se debe seguir para reetiquetar los archivos con las nuevas etiquetas que sustituyen a los nombres de atributos originales, y que facilitarán el proceso de ingreso de los datos a la base de datos. En este diagrama se puede ver que una vez que se ha seleccionado el proceso de reetiquetado de archivos el sistema responde mostrando un formulario, en el cual el usuario debe escribir el nombre del archivo que será procesado; el proceso que se realiza a este archivo es: leer su contenido línea por línea, si la línea contiene el símbolo “:” eso será un indicativo de que se trata de una etiqueta por lo cual se hace una búsqueda de la etiqueta en la relación que almacena el catálogo de autoridades de códigos, se obtiene el código correspondiente a la etiqueta y se escribe este código, junto con los datos a él asociados en el archivo de datos reetiquetado. Como resultado se obtiene un archivo reetiquetado y un mensaje indicando el total de registros procesados. En la Tabla 4.4 se muestra un fragmento de este tipo de archivos.

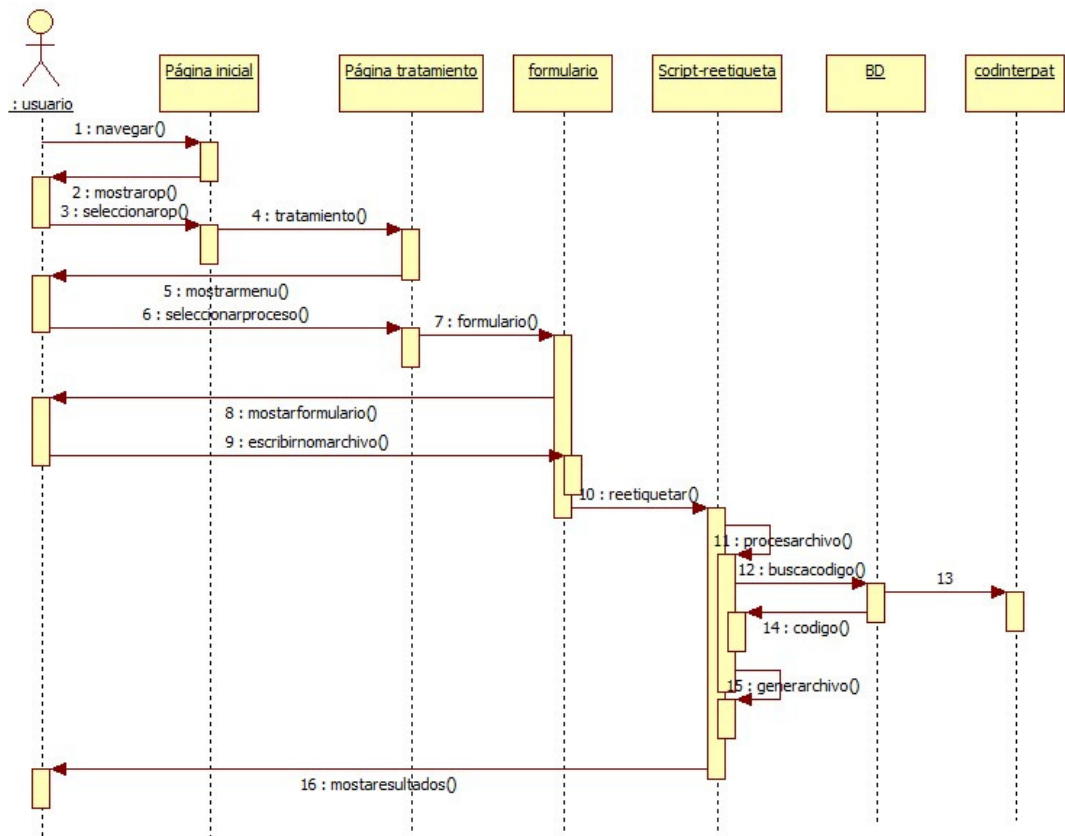


Figura 4.7: Diagrama de secuencia de reetiquetado.

Tabla 4.4: Fragmento de un archivo de tipo texto reetiquetado.

PT P200500164
TI PROCESOS "VIA HUMEDA" PARA LA PRODUCCION DE FRAGMENTOS Y PASTILLAS FUMIGENAS Y LACRIMOGENAS RESINIFICADAS CON DISMINUCION DE GASES NO ANTROPOGENICOS TIPO "CFC" (CLOROFLUOROCARBONADOS)
TP P
FS 20050128
FP 20060901
FI 20060901
NS P200500164
FC 20090223
FE 20090401
NP 2258916
CP A1 B2
SO FALKEN, S.A.
DS DIAMANTE N. 7, N.2.,MADRID 28021
PS 28
NA ES
IN BARDISA JORDA,JOSE IGNACIO
NI ES
CE C06B29/02 C06D3/00 C06D7/00
CS C06B29/02 (2006.01) C06D3/00 (2006.01) C06D7/00 (2006.01)
IE US3391036 US233 US4032374 GB1520601 US6436210 US5154782 US5522320
CA X,X,X,X,X,X,X
RE Procesos "vía húmeda" para la producción de fragmentos y pastillas fumígenas y lacrimógenas resinificadas con disminución de gases no antropogénicos tipo "CFC" (clorofluorocarbonados).#La presente invención se refiere a una mezcla pirotécnica y a su procedimiento de fabricación manual, semi automático y automático de fragmentos y pastillas de mezclas fumígenas y/o lacrimógenas por un proceso de vía húmeda mediante gelificación y por polimerización o resinificación de las mezclas de ingredientes pirotécnicos.

4.2.1.3. Ingresar datos *Interpat*

Otra funcionalidad requerida en el módulo de carga de datos es la relacionada con la adición de los datos de la base de datos *Interpat* al modelo de integración aquí propuesto, esta funcionalidad es de gran importancia debido a que, como se ha mencionado, el punto de partida son los registros de patentes de la oficina local. En este apartado se describen los requerimientos funcionales que se deben cumplir para satisfacer esta funcionalidad.

Precondiciones. Los archivos de *Interpat* en formato texto deben estar reetiquetados con las etiquetas del catálogo de autoridades creado para este fin.

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Solicitar nombre del archivo *Interpat* que será ingresado.
- v. Para cada registro de patente contenido en este archivo se deben ingresar los datos en la correspondiente relación del repositorio de datos.
- vi. Mostrar el número de registros ingresados a la base de datos.
- vii. Mostrar el número de registros duplicados encontrados en el archivo procesado.
- viii. Condición de error: si no se encuentra el archivo indicado, se envía mensaje de error.

En la Figura 4.8 se muestra el diagrama de secuencia para ingresar los datos *Interpat*. Para realizar la inserción de los datos, se muestra un formulario en el cual se debe digitar el nombre del archivo reetiquetado que será procesado, después de escribir este nombre se hace un llamado al *script* que se encarga de procesar los registros contenidos en el archivo e insertarlos en la relación *patenteoepm*. Al finalizar este procesamiento de los datos se muestra un mensaje indicando el número de registros que fueron ingresados.

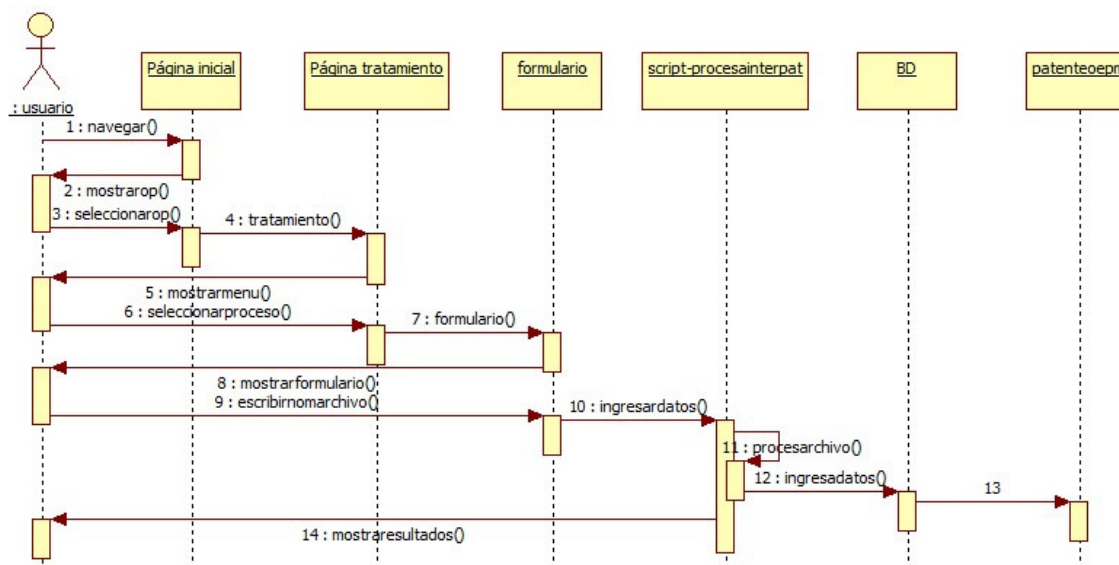


Figura 4.8: Diagrama de secuencia para procesar datos *Interpat*.

4.2.1.4. Generación de lista con números de patentes *Interpat*

Una vez que se han ingresado los datos extraídos de la base de datos *Interpat*, es el momento para realizar la búsqueda de esas patentes en la base de datos *Derwent Innovation Index*, para ello se genera una lista con los números de patentes, esta lista o listas de números serán utilizados por el usuario en la interfase de búsqueda proporcionada por *Derwent*.

Precondiciones. La base de datos de patentes debe tener almacenados los documentos de patentes descargados de la base de datos *Interpat*.

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Generar las listas con los números de patentes, mostrarlas en el formato de búsqueda utilizado en la interfaz de la base de datos *Derwent*.

Al igual que en los casos anteriores, del menú mostrado por la opción tratamiento de datos, el usuario selecciona la opción “Generación de listas con números de patentes *Interpat*”, esta opción hace una llamada al *script* encargado de generar estas listas, para ello se realiza una consulta en la relación *patenteoepm* y se seleccionan los números de patentes. También, es importante mencionar que los números de patente obtenidos a través de la base de datos *Interpat* no tienen como prefijo el código de dos letras del país, así, el *script* que genera estas listas antepone el código **ES** a cada uno de los números de patente que serán buscados en la base de datos *Derwent Innovation Index*. En el anexo se muestra una lista completa de los códigos de dos letras utilizados en el formato del número de las patentes. El *script* implementado para generar estas listas utiliza un parámetro de 50 números de patente por línea de búsqueda, este parámetro no es una limitante y puede ser cambiado para que se generen listas con un mayor número de elementos. En la Figura 4.9 se muestra el diagrama de secuencia correspondiente a este proceso.

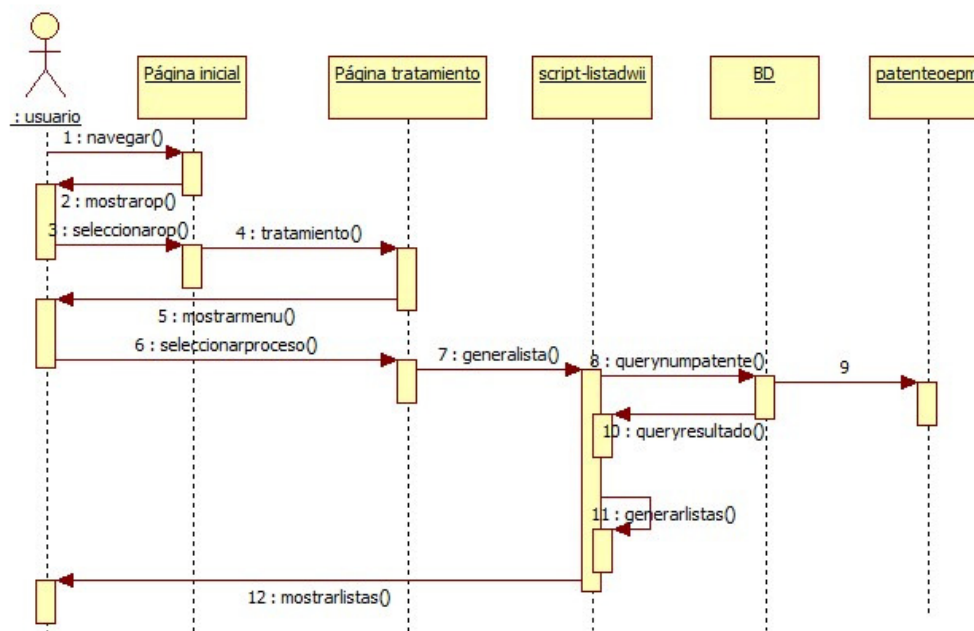


Figura 4.9: Diagrama de secuencia para generar lista de números de patentes.

Una vez que se han generado estas listas o líneas con números de patentes, el usuario deberá utilizar la interfaz de búsqueda de *Derwent Innovation Index*, buscar y

descargar estos registros. Cuando se realiza el proceso de descarga, la interfaz de *Derwent Innovation Index* muestra una lista con los datos que se pueden descargar de los registros seleccionados, se deberán seleccionar todos los datos con excepción de los dibujos. En las Figuras 4.10 y 4.11 se muestran una pantalla con una lista de líneas de búsqueda, y una pantalla con la interfaz de búsqueda de *Derwent Innovation Index*.

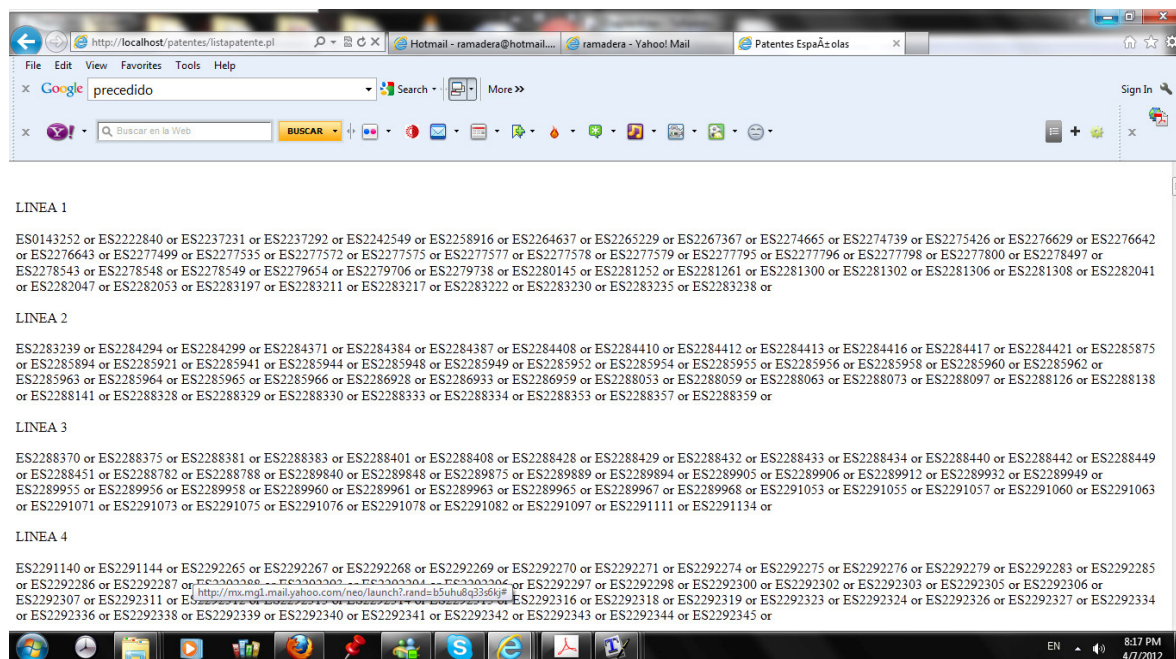


Figura 4.10: Lista de líneas de búsqueda de números de patentes.

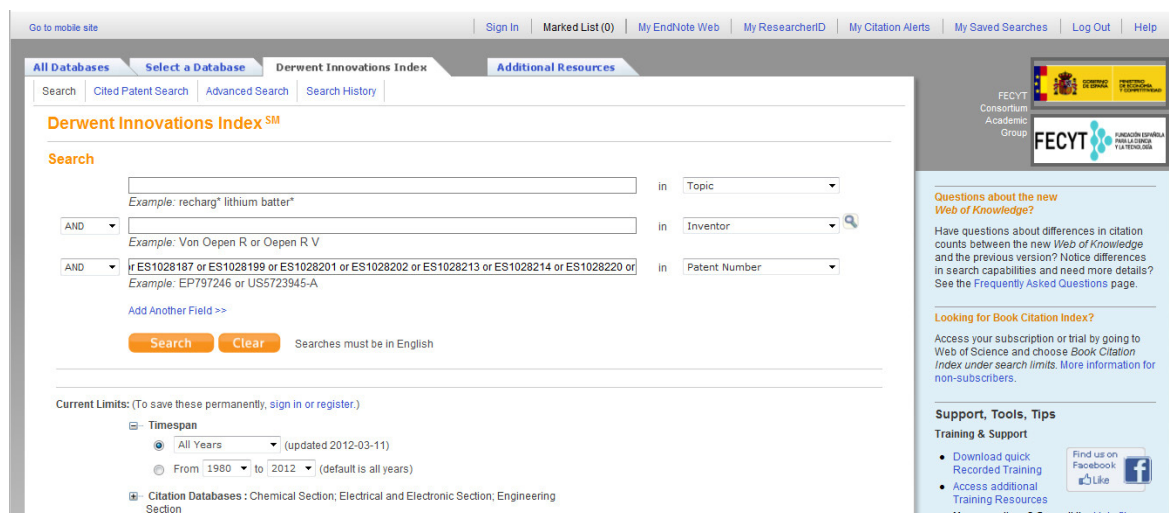


Figura 4.11: Interfaz de búsqueda de *Derwent Innovation Index*.

4.2.1.5. Ingresar datos *Derwent Innovation Index*

Como se ha visto, para satisfacer el requerimiento funcional de ingresar los registros de patentes de la base de datos *Derwent Innovation Index*, una parte la debe hacer el usuario en forma independiente, es decir, mediante la búsqueda y descargar de los registros de patentes a través de la interfaz de búsqueda de esta base de datos. Cuando se han descargado los registros de las patentes se completa esta actividad a través del tratamiento de datos que aquí se está proponiendo. En la Tabla 4.5 se muestra un fragmento de los archivos con registros de patentes de la base de datos *Derwent Innovation Index*.

Tabla 4.5: Fragmento de un archivo de la base de datos *Derwent Innovation Index*.

```

FN ISI Export Format
VR 1.0
PT P
PN ES2156058-A1; ES2156058-B1
TI Peripheral device for CD-ROM with multi-media capacity involves reading-printing unit with
optical unit control in circuitry, signal generation and processing systems necessary for unit to have
multi-media capacities.
AE UNIV DALACANT (UYDA-Non-standard)
GA 2003--002328
DC T03 (Data Recording)
MC T03-B06; T03-N01
IP G06F-003/16; G11B-031/00
PD ES2156058-A1 01 Jun 2001 G06F-003/16 200301 Pages: 1
    ES2156058-B1 01 Mar 2002 200301
AD ES2156058-A1 ES002153 16 Oct 1998
PI ES002153 16 Oct 1998
CP ES2156058-A1
    WO9629641-A1 ELONEX TECHNOLOGIES INC (ELON-Non-standard) KIKINIS D
CR ES2156058-A1
    BASE DE DATOS WPI en EPOQUE, Derwent Publications Ltd. (Londres: GB) semana 199321
& RESEARCH DISCLOSURE Vol. 348 Nr. 006 (ANONIMO) 10.04.1993
UT DIIDW:2003002328
ER
    
```

A continuación se describirán los requerimientos necesarios para completar esta tarea.

Precondiciones. Los archivos de *Derwent* deben estar en el formato *plain text*.

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.

- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Solicitar nombre del archivo *Derwent* que será procesado.
- v. Para cada registro de familia de patente contenido en este archivo se debe ingresar en la base de datos los siguientes datos en las relaciones correspondientes.
- vi. Mostrar el número de registros ingresados a la base de datos.
- vii. Mostrar el número de registros duplicados encontrados en el archivo procesado.
- viii. Condición de error: si no se encuentra el archivo indicado, se envía mensaje de error.

En la Figura 4.12 se muestra el diagrama de secuencia para ingresar los datos *Interpat*. En este diagrama se puede observar que una vez que se ha seleccionado el archivo con datos, de la base de datos *Derwent Innovation Index*, se inicia un proceso que consiste en leer cada uno de los registros de este archivo, extraer, datos generales de la familia de patentes, los números de patente de la familia, las patentes referenciadas, la clasificación internacional de patentes, los nombres de los solicitantes, los nombres de los inventores. Estos datos se almacenan en la relación correspondiente a cada una de ellos. En relación con los datos de la clasificación internacional de patentes, además de almacenarlos en la relación que les corresponde, se hace un procesamiento que consiste en obtener el primer carácter de la clasificación y almacenarlo en la relación *ipcdwii1*, debido a que es aquí donde se tienen los datos de clasificación a nivel clase. El mismo proceso se sigue para obtener la clasificación a nivel subgrupo, en este caso se obtienen los primeros 3 caracteres de la clasificación y se almacenan en la relación *ipcdwii2*.

Otro procesamiento presente en el diagrama es la generación de la interrelación existente entre los datos de las bases de datos *Interpat* y *Derwent Innovation Index*, para generar esta interrelación se buscan los números de patente en la relación *patenteoepm*, obteniéndose el número de patente, y el año de concesión. Los datos obtenidos de este proceso se almacenan en la relación *patentedwoepm*, posteriormente se hace la actualización del año de concesión en los registros insertados al inicio del proceso de “Ingreso de datos *Derwent Innovation Index*”.

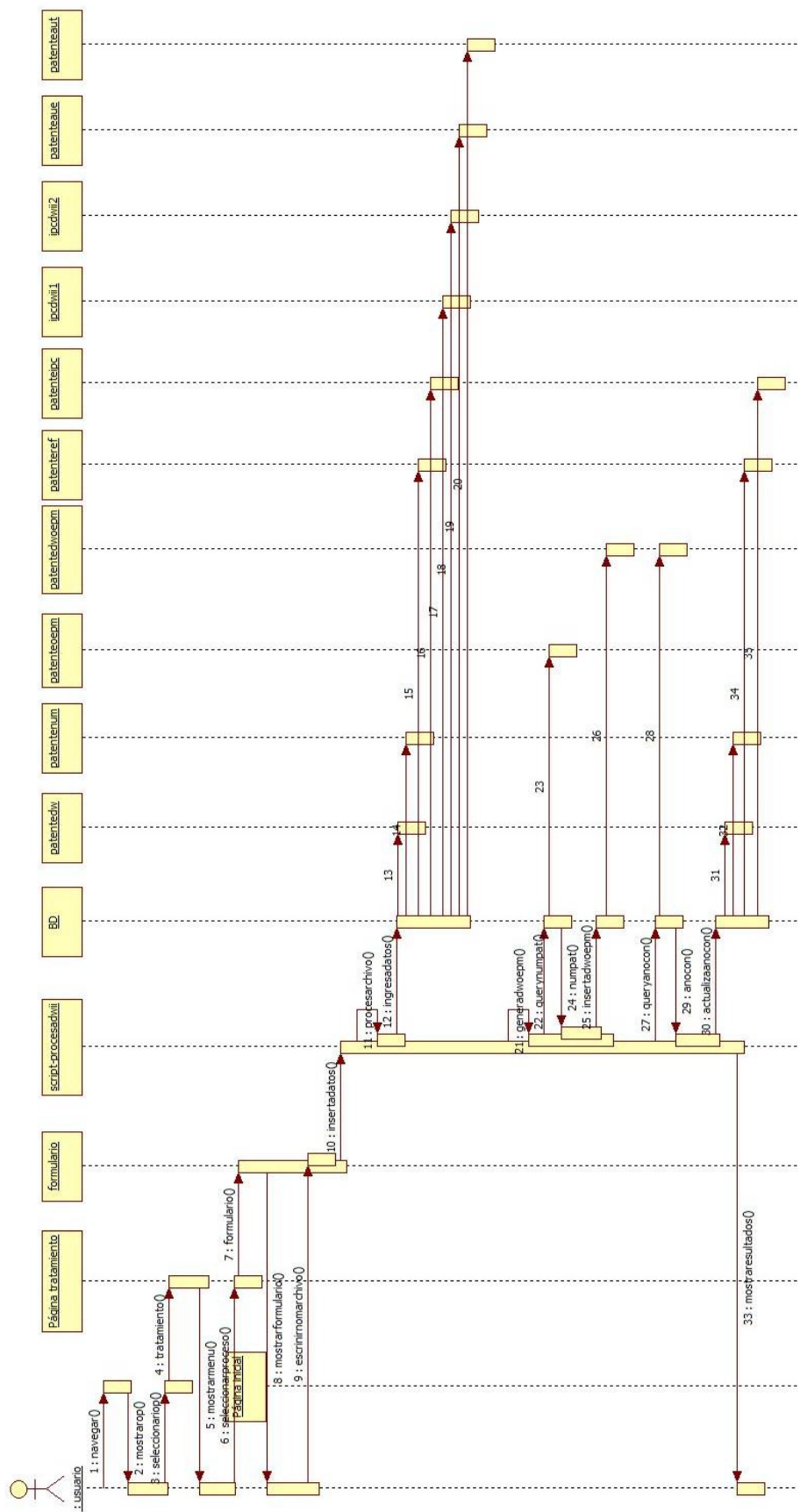


Figura 4.12: Diagrama de secuencia para procesar archivos *Derwent Innovation Index*.

4.2.1.6. Generación de lista con números de patentes referenciadas

El proceso para generar la lista de patentes referenciadas es similar al descrito en la sección *Generación de lista con números de patentes Interpat*. En este caso los números de patentes corresponden a los números de patentes referenciadas por los inventores y examinadores de los registros descargados de la base de datos *Derwent Innovation Index*.

Precondiciones: i) La base de datos de patentes debe tener almacenados los documentos de patentes descargados de la base de datos *Derwent Innovation Index*; ii) Las listas de términos de búsqueda deben ser de 10 elementos.

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Generar enlaces de búsqueda, de 10 números de patentes referenciadas, a la base de datos *Worldwide-espacenet*.

El portal *espacenet* limita la búsqueda a 10 términos por cada campo proporcionado en la interfaz de búsqueda de la base de datos *Worldwide-espacenet*. Lo anterior implicó que se restringieran los enlaces generados en el requerimiento 4. En la Figura 4.13 se muestra el diagrama de secuencia para generar la lista de patentes referenciadas, el proceso que se sigue para generar las listas de las patentes referenciadas es similar al descrito para generar las listas con los números de patentes *Interpat*. En este caso los números de patentes referenciadas se toman de la relación *pateneref*. Cada línea de esta lista tiene el siguiente formato:

```
<a  
href=http://v3.espacenet.com/searchResults?DB=EPODOC&submitted=true&locale=en_EP&ST=advanced&PN=147600+OR+158630+OR+2000662+OR+516277+OR+671135+OR+A+OR+A+OR+A+OR+AARESTRUP+OR+AARNIO&compact=true>1</a>http://v3.espacenet.com/searchResults?DB=EPODOC&submitted=true&locale=en_EP&ST=advanced&PN=147600+OR+158630+OR+2000662+OR+516277+OR+671135+OR+A+OR+A+OR+A+OR+AARESTRUP+OR+AARNIO&compact=trac<br>
```

Al generar este tipo de líneas de búsqueda, el usuario buscará de forma directa las patentes referenciadas en la base de datos *Worldwide-espacenet*. En la Figura 4.14 se puede observar una página *Web* que contiene estas listas.

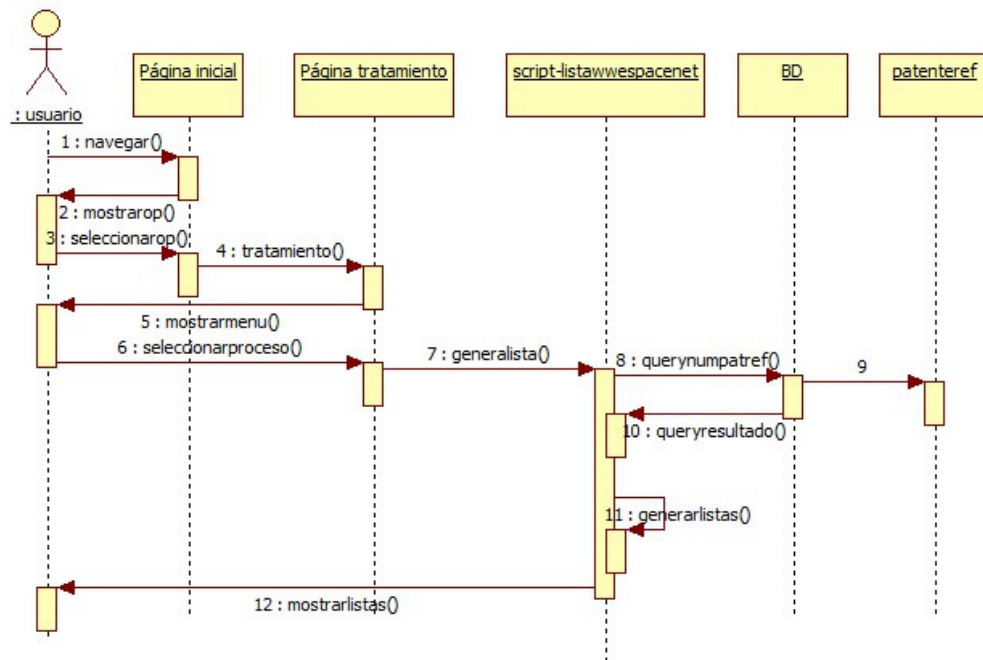


Figura 4.13: Diagrama de secuencia para generar lista de patentes referenciadas.

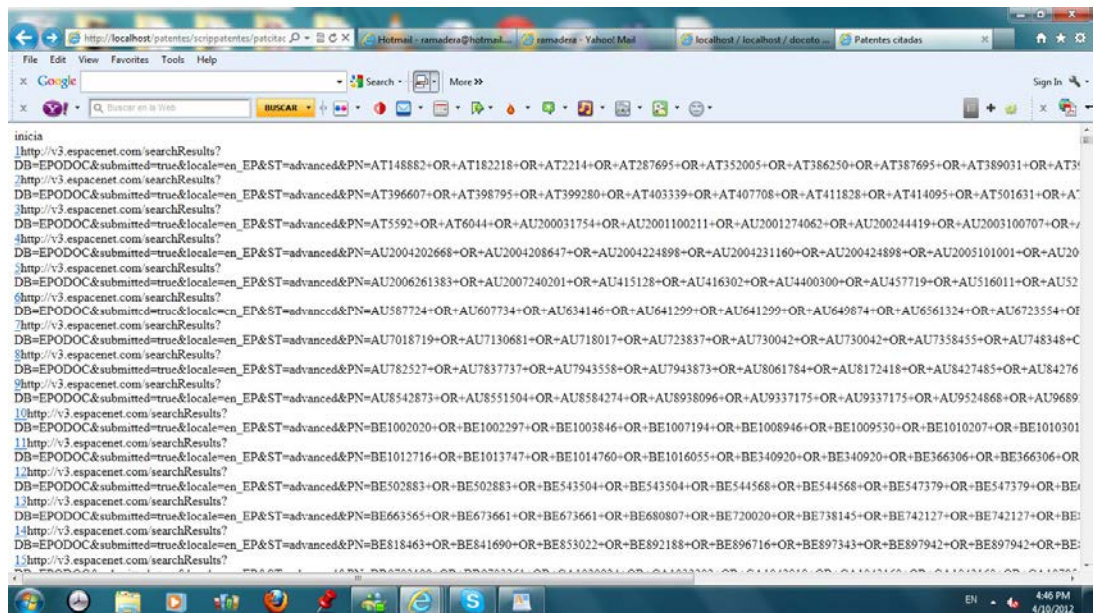


Figura 4.14: Lista de líneas de búsqueda de patentes referenciadas.

4.2.1.7. Ingresar datos Worldwide-espacenet

Al igual que en el caso de ingresar los registros de patentes de la base de datos *Derwent Innovation Index*, para procesamiento de los registros de la base de datos *Worldwide-espacenet* el usuario debe realizar en forma independiente la descarga de los registros de patentes a través de la interfaz de búsqueda de esta base de datos. En la Figura 4.15 se muestra un fragmento de los archivos con registros de patentes citadas de la base de datos *Worldwide-espacenet*, y a continuación se describen los requerimientos funcionales necesarios para completar esta actividad.

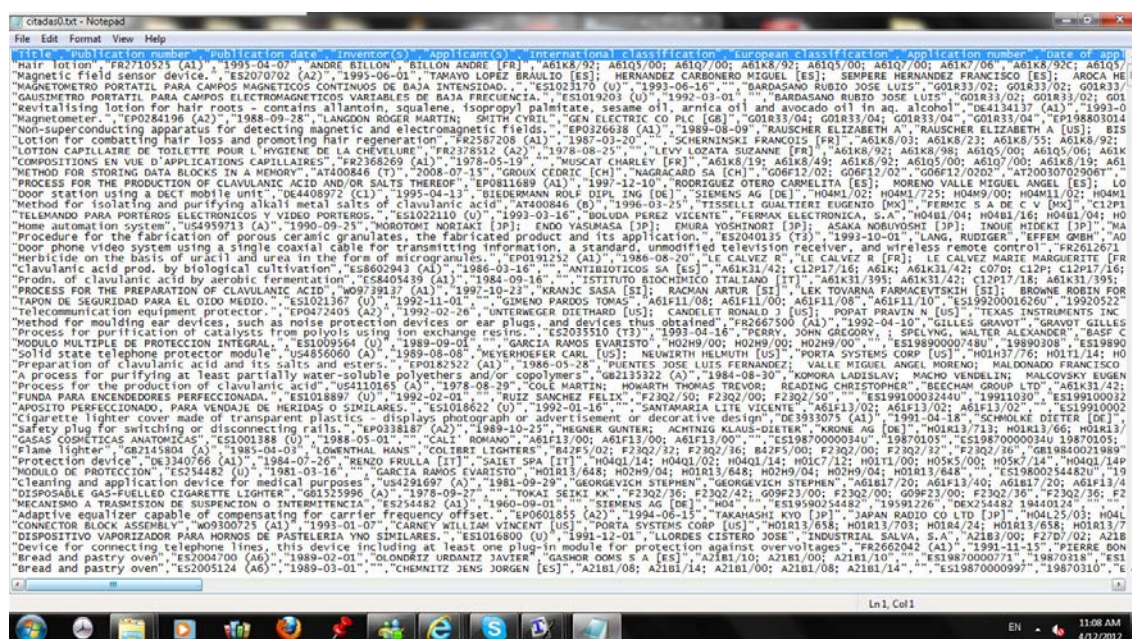


Figura 4.15: Fragmento de patentes citadas, formato csv.

Precondiciones. i) Los archivos con patentes referenciadas deben estar en formato CSV; ii) Los archivos deben contener los datos mostrados en el Tabla 3.3.

- El usuario entra a la página inicial del sistema.
- Mostrar menú de opciones generales disponibles.
- Mostrar procesos correspondientes al tratamiento de datos.
- Solicitar el nombre del archivo CSV que será procesado.
- Para cada registro de patente contenido en este archivo se deben ingresar los datos en la correspondiente relación del repositorio de datos.

- vi. Mostrar el número de registros ingresados a la base de datos.
- vii. Mostrar el número de registros duplicados encontrados en el archivo procesado.
- viii. Condición de error: si no se encuentra el archivo indicado, se envía mensaje de error.

En la Figura 4.16 se muestra el diagrama de secuencia para ingresar los datos *Worldwide-espacenet*. Para realizar la inserción de los datos, se sigue un proceso similar al utilizado para insertar los registros de la base de datos *Interpat*, en el caso de los registros de patentes referenciadas los datos se almacenan en la relación *patentespacenet*. Al finalizar este procesamiento se muestra un mensaje indicando el número de registros que fueron ingresados.

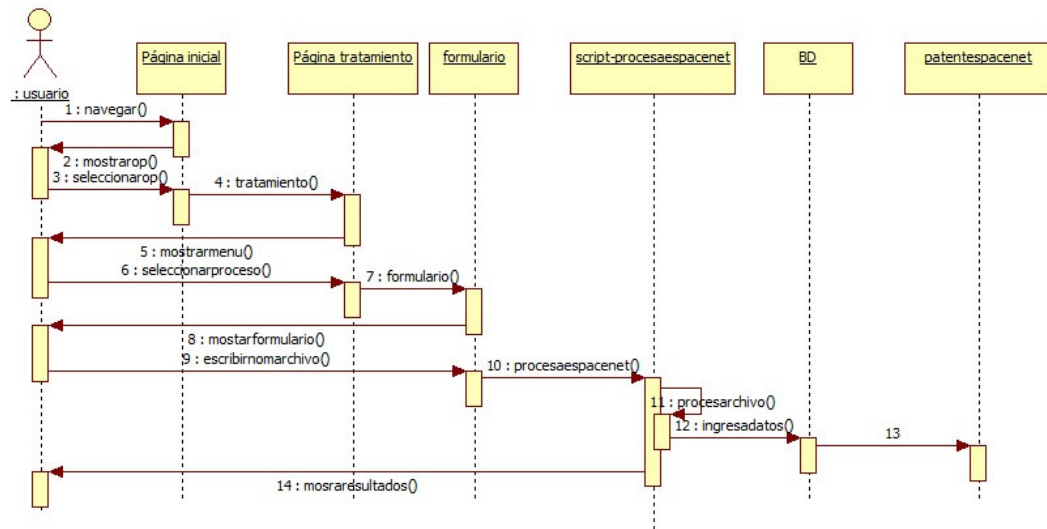


Figura 4.16: Diagrama de secuencia para ingresar patentes referenciadas.

4.2.1.8. Procesar patentes referenciadas

En el epígrafe anterior se describió el proceso para insertar los datos de las patentes referenciadas descargadas de la base de datos *Worldwide-espacenet*. Los registros insertados contienen todos los datos proporcionados por esta base de datos. Dado que los datos se utilizan de la base *Worldwide-espacenet*, estos son: el número de patente, la nacionalidad del solicitante, el año de publicación del documento y la clasificación

internacional de patentes. Es preciso realizar un procesamiento para extraerlos de su forma original, para ello es necesario que se cumplan los requerimientos siguientes:

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Para cada registro de patente contenido en la relación *patentespacenet* extraer el número de patente, la nacionalidad del solicitante, el año de publicación del documento, y la clasificación internacional de patentes.
- v. Almacenar los datos extraídos en la relación correspondiente del repositorio de datos.
- vi. Mostrar el número de registros ingresados a la base de datos.

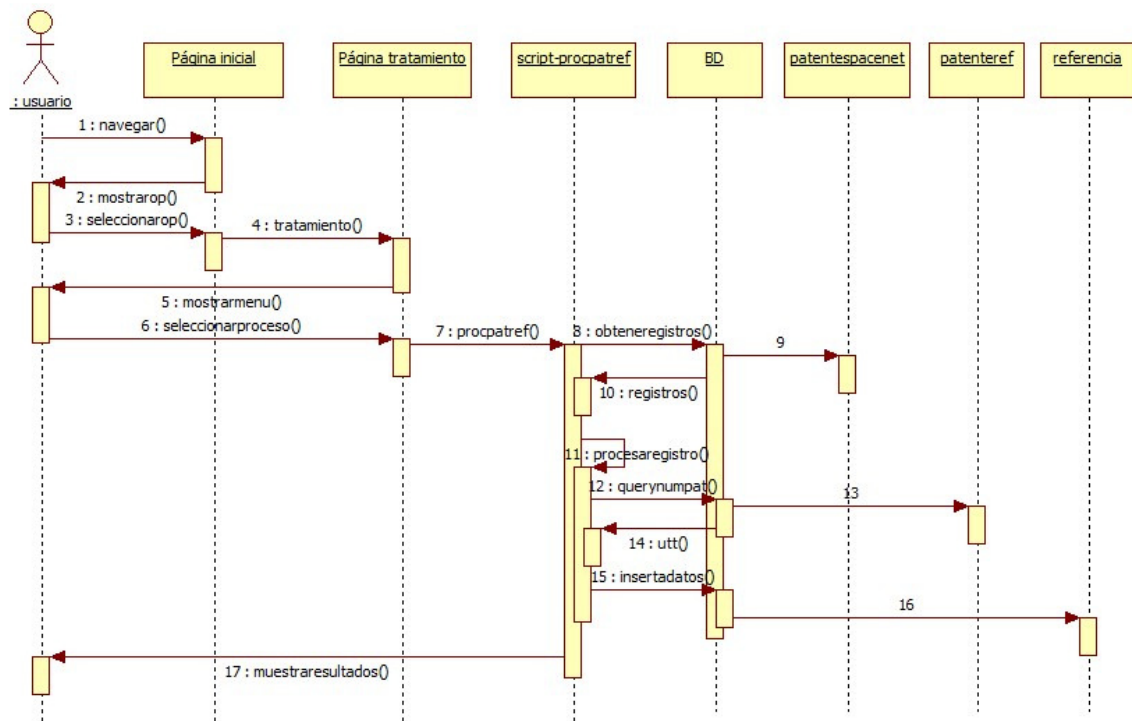


Figura 4.17: Diagrama de secuencia para ingresar patentes referenciadas.

En la Figura 4.17 se muestra el diagrama de secuencia para el procesamiento de los datos de las patentes referenciadas. En este proceso se extrae el número de la patente, de la fecha de publicación se obtiene el año en formato de 4 dígitos, para cada

solicitante, se obtiene su nacionalidad tomando el código de dos letra que se encuentra entre corchetes, la clasificación se separan en partes unitarias (recordar que están separadas por un ;), se obtiene el número de identificación único de la familia de patentes citante, el número de patente citante, y finalmente se almacenan estos datos en la relación que contiene los datos necesarios de las patentes referenciadas.

4.2.2. Normalización

Como se sabe, uno de los principales problemas al realizar estudios métricos de la información es la falta de normalización de los nombre de los autores y direcciones de afiliación, el análisis de las patentes tampoco escapa a esta problemática, para solventar este problema en la normalización de solicitantes se desarrolló un módulo que permite normalizar los nombre y su clasificación por sectores a través de una interfaz gráfica. Para realizar el proceso de normalización de los solicitantes de patentes, los datos extraídos de las bases de datos antes mencionadas son almacenados en una base de datos generada con el esquema preliminar del *Data Warehouse*.

El propósito de desarrollar un módulo que pueda ser utilizado como herramienta de apoyo para normalizar los nombres de los solicitantes de patentes y codificarlos por sectores.

4.2.2.1 Normalizar nombre del solicitante

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Mostrar todos los registros de la relación *patenteinstitucion* que no tengan datos de nombre de solicitante normalizado.
- v. Mostar nombres normalizados.
- vi. Actualizar el identificador del nombre del solicitante normalizado.
- vii. Regresar al paso 4.

El proceso de normalización de nombres de solicitantes se lleva a cabo en la relación *patenteinstitucion*. Después de seleccionar esta actividad dentro del menú de opciones de tratamiento de datos se ejecuta el *script* de normalización, éste busca todos los registros que no tengan dato en el atributo *solnormalizado*, para cada uno de los registros encontrados busca, en la misma relación, el nombre del solicitantes en el atributo solicitante pero cuyo contenido en el atributo *solnormalizado* no este en blanco, si lo encuentra hace una actualización del identificador del solicitante (idnombre), del tipo de sector asignado (idtipoinstitucion), y actualiza el valor del atributo *solnormalizado*. Al finalizar el proceso se muestran la cantidad de registros normalizados, en la Figura 4.18 se puede observar este procedimiento.

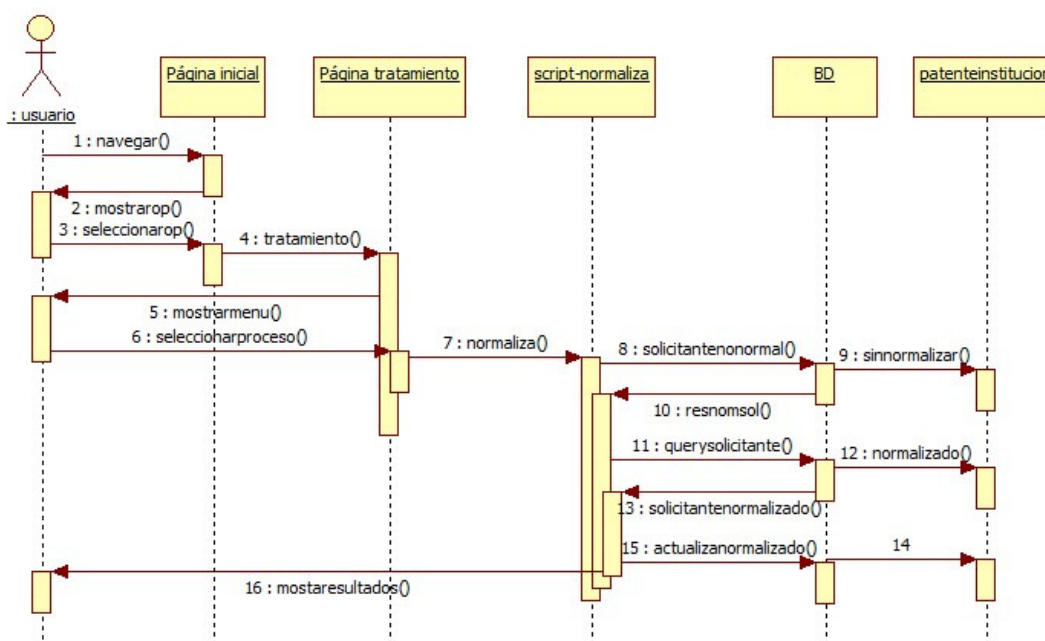


Figura 4.18: Diagrama de secuencia para normalizar nombre de solicitantes.

4.2.2.2 Codificación de solicitante

La finalidad de codificar los solicitantes es la de agruparlos por sectores productivos. Los sectores que se utilizan para este tipo de clasificación son: Universidades, Consejo Superior de Investigaciones Científicas (CSIC), Empresas, Particulares, Organismos Públicos de Investigación (OPI), Universidades extranjeras, y Otros.

Estos tipos de sectores se encuentran almacenados en el catálogo de autoridad de sectores. Los requerimientos para completar esta actividad se listan a continuación:

- i. El usuario entra a la página inicial del sistema.
- ii. Mostrar menú de opciones generales disponibles.
- iii. Mostrar procesos correspondientes al tratamiento de datos.
- iv. Mostrar todos los registro contenidos en la relación *patenteinstitucion* que tenga código de sector 0.
- v. Mostar catálogo de autoridad de sectores.
- vi. Actualizar el código del sector asignado al nombre del solicitante.
- vii. Regresar al paso 4.

El proceso para codificar el sector asignado al solicitante inicia una vez que el usuario ha seleccionado realizar esta actividad, el *script* de codificación hace una búsqueda de todos aquellos solicitantes que aún no tengan asignado un sector (el código asignado a estos solicitantes es 0), después se obtienen todos los sectores disponibles del catálogo de autoridad de sectores (relación *tipoinstitucion*). Una vez que se tienen todos los datos se muestran en pantalla los nombres de los solicitantes y los sectores disponibles. El usuario deberá seleccionar el solicitante y su sector correspondiente. Al aceptar los valores seleccionados se hace una actualización del código de sector asignado al solicitante, el proceso se repetir el proceso mientras haya solicitantes sin sector asignado. En la Figura 4.19 se muestra el diagrama de secuencia de esta actividad.

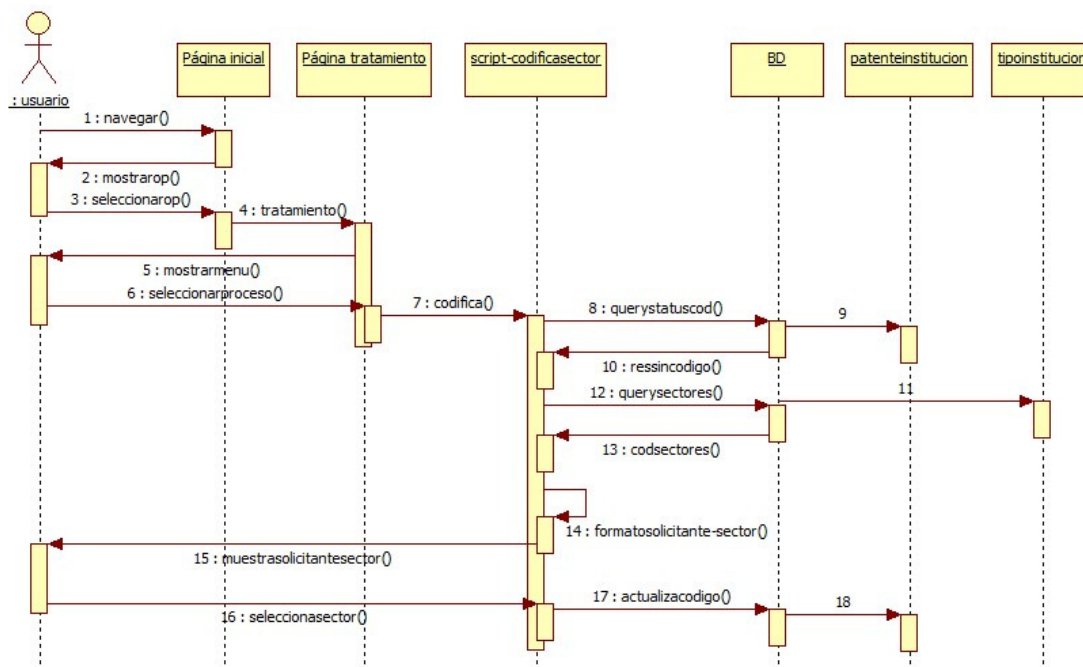


Figura 4.19: Diagrama de secuencia para codificar sector de solicitantes.

4.2.3. Generación de reporte

Otros de los mecanismos necesarios para llevar a cabo la explotación de un repositorio de datos, y que constituyen una de las tareas es la metodología KDD, son las herramientas para la obtención de reportes que permitirán realizar el análisis de los datos y la visualización de los mismos. En esta sección se definirán los requerimientos funcionales que permiten desempeñar estas actividades, así como los diagramas de secuencia ligados a éstos.

El propósito de desarrollar un módulo que permita generar reportes de los datos almacenados en el repositorio, los cuales puedan ser utilizados para obtener indicadores de patentes, así como para ser utilizados en herramientas de análisis y visualización de datos.

Los reportes que se generan en este módulo, y que son utilizados para obtener los indicadores de patentes y realizar el análisis de estos son los siguientes:

- i. Generar un reporte que muestre la evolución de las patentes concedidas.
- ii. Generar reporte que permita realizar análisis de la evolución de las patentes concedidas por sector.
- iii. Generar reportes que permitan conocer la evolución de las patentes concedidas por clases y subgrupos a segundo nivel.
- iv. Generar reportes que permitan conocer la distribución de las patentes concedidas por sector, clases, y subgrupos a segundo nivel.
- v. Generar reportes que permitan conocer la evolución de las patentes prioritarias concedidas.
- vi. Generar reportes que permitan conocer la evolución, la distribución nacional, distribución por clases y subgrupo a segundo nivel de las patentes referenciadas por las patentes concedidas.
- vii. Generar reportes que permitan conocer la evolución, distribución por temática, países de protección de las familias de patentes concedidas.
- viii. Generar reportes que permitan analizar la evolución, y temática a primero y segundo nivel las patentes triádicas.

Los requerimientos funcionales que deben ser cumplidos para generar cada uno de los reportes antes mencionados son:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte solicitado.

En los requerimientos funcionales enumerados arriba se puede observar que los reportes que sirven como base para obtener indicadores de patentes tienen en común una sección que corresponde a la interacción que existe entre el usuario o analista y la interfaz del módulo para solicitar los reportes. En la Figura 4.20 se muestra el diagrama de secuencia que es común en todos estos reportes.

En el diagrama de secuencia de la Figura 4.20 se observa cómo el usuario o analista de datos entra en la página inicial del sistema que le permite realizar operaciones en el repositorio de datos. Esta primera página es una página *Web* generada por un *script*, en ella se muestran dos opciones que corresponden a las actividades de tratamiento de datos y generación de reportes (ver Figura 4.3). En el diagrama de secuencia de la Figura 4.20 el analista de datos selecciona la opción de generación de reportes, al seleccionar esta opción el *script* hace un llamado a un segundo *script*. Este segundo *script* genera una página *Web* que contiene un formulario en el cual el analista selecciona los parámetros que son utilizados en la consulta (ver Figura 4.21). Una vez aceptados estos parámetros se hace un llamado al *script* que muestra un menú con una lista de reportes cuyos resultados podrán ser utilizados para la obtención de indicadores de patentes; al seleccionar un reporte de esta lista se hace un enlace al *script* que genera los resultados solicitados. En los epígrafes siguientes se describen los detalles específicos de cada uno de los *script* encargados de mostrar los resultados de la lista de reportes disponibles.

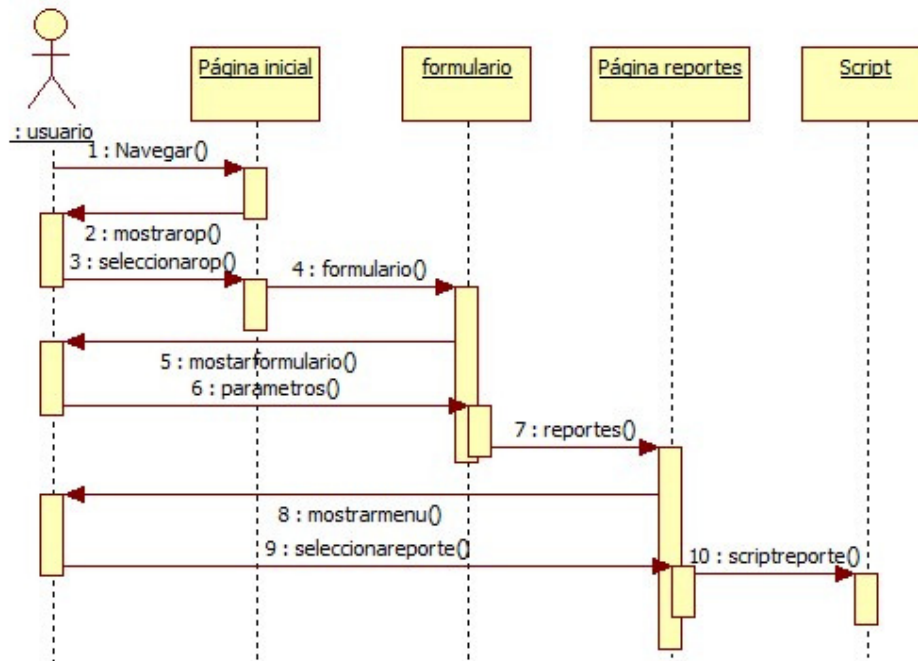


Figura 4.20: Diagrama de secuencia general.

The screenshot shows a web browser window with the address bar displaying 'http://localhost:patentes/scrippat'. The browser's title bar says 'Patentes Derwent'. The page content is a form with the following fields:

- Tipo de documento: Patentes (OEPM) (dropdown menu)
- Comunidad: (empty text box)
- Percentiles: (empty text box)
- Umbral: (empty text box)
- Año inicial: 1996 (dropdown menu)
- Año final: 2007 (dropdown menu)
- aceptar (button)

Figura 4.21: Página de parámetros.

4.2.3.1. Evolución general de patentes

Este reporte proporciona la distribución del número de patentes que se han concedido anualmente en el periodo seleccionado, en la Figura 4.22 se muestra el diagrama de secuencia para generar el reporte. En el formulario de parámetros de consulta mostrado en la Figura 4.21 se debe seleccionar la opción Patentes OEPM, y periodo. Al completar y aceptar los valores digitados en el formulario se realiza una invocación al script encargado de obtener los datos del repositorio de datos, éstos se obtienen de la relación *patentoeopm*, son contabilizadas las patentes concedidas dentro del periodo solicitado, se agrupan por año de concesión, los datos son mostrados en el navegador y se separan por el símbolo (|). En este caso no se tiene en cuenta el valor del umbral. Con estos resultados el analista tiene la posibilidad de utilizarlos en alguna herramienta de graficación o simplemente darles el formato de presentación adecuado. En la Figura 4.23 se muestra una pantalla con datos de la evolución general generada.

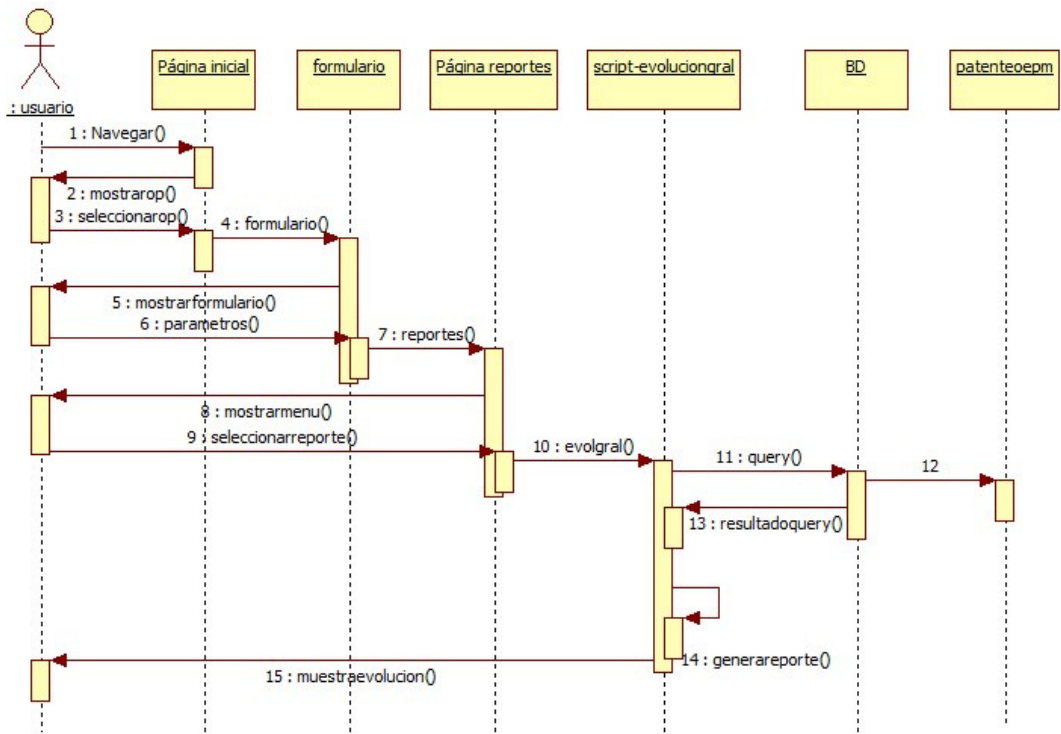


Figura 4.22: Diagrama de secuencia para generar evolución general de patentes.

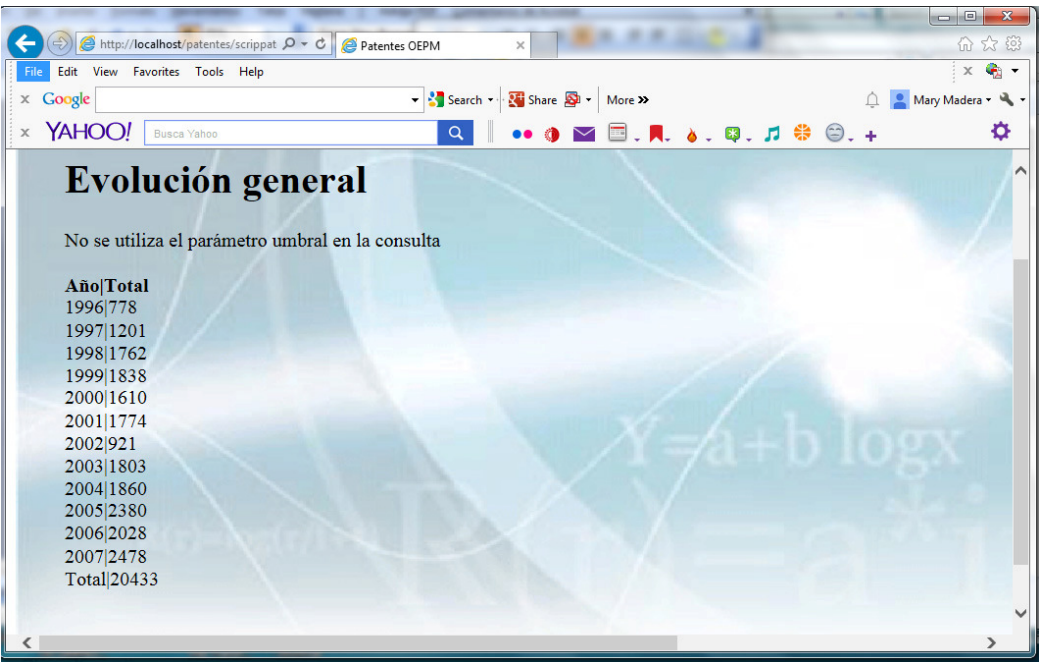


Figura 4.23: Pantalla con reporte de evolución general de patentes.

4.2.3.2 Distribución general y evolución por sectores

El este reporte se presentan dos grupos de datos, en el primer grupo se muestra la distribución general por sector, y en el segundo grupo de datos se tiene la evolución por sector.

Los requerimientos funcionales que deben cumplirse para generar la distribución y evolución por sectores de las patentes concedidas en el periodo seleccionado son los siguientes:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución general y evolución por sectores.
- v. Mostrar los resultados

Para generar este reporte el analista debe llenar el formulario con la opción Patentes OEMP, comunidad, umbral, y periodo, al proporcionar y aceptar estos valores se ejecuta el *script* que genera un reporte con los datos solicitados. Este script utiliza las relaciones *patenteoepm*, *patenteinstitucion*, y *tipoinstitucion*. Para generar la primera parte del reporte (la distribución general por sector) se cuentan las patentes concedidas en el periodo seleccionado de la comunidad indicada, agrupadas por sector, no se toma en cuenta el umbral seleccionado, y si no se indicó ningún valor en el parámetro comunidad se toma en cuenta todos los registros. Si en una patente hay más de un sector se le asigna un punto a cada uno de los sectores solicitantes.

En el segundo grupo de datos se realiza un procesamiento de datos similar al descrito anteriormente, pero en este caso la agrupación se hace por año y por sector. Los datos son formateados y mostrados en pantalla para que el analista de datos realice la graficación y análisis pertinentes. En la Figura 4.24 se muestra el diagrama de secuencia.

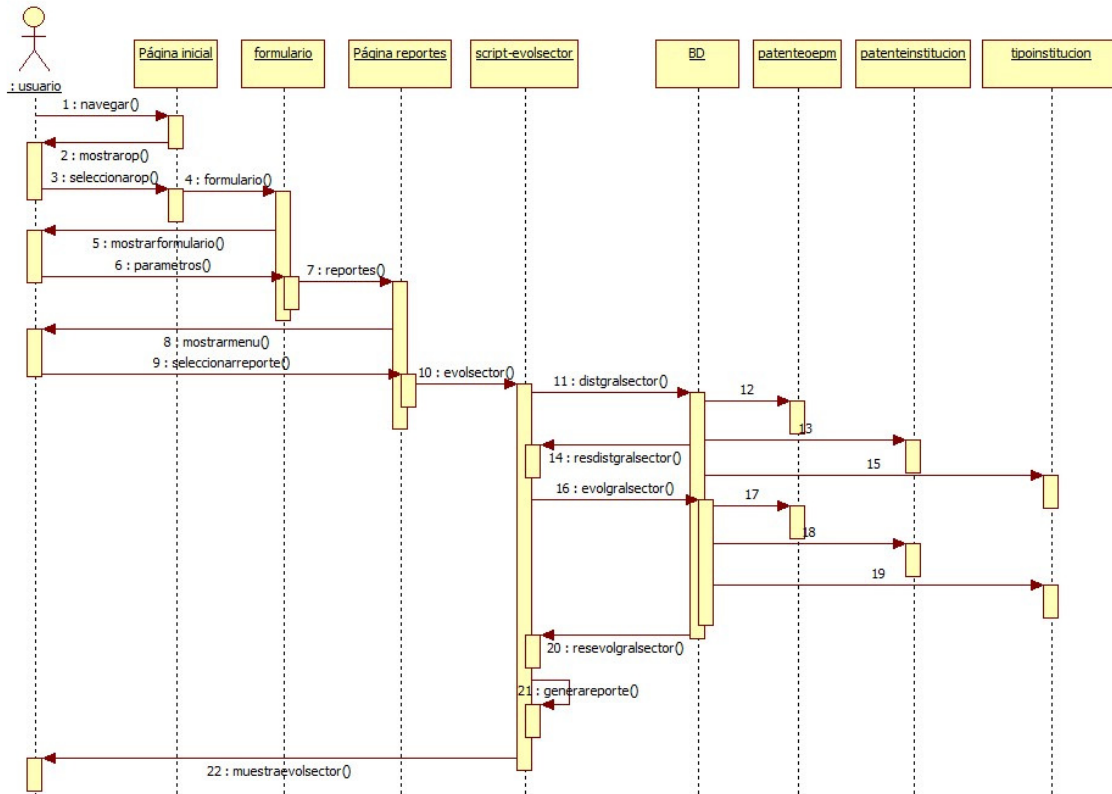


Figura 4.24: Diagrama de secuencia para generar evolución general por sector.

4.2.3.3. Distribución general y evolución por temática a nivel clase

En este tipo de reporte se obtienen tablas de datos que muestran la distribución por área temática a nivel clase y evolución de las patentes, es decir, la distribución y evolución de la clasificación internacional de las patentes concedidas tomando en cuenta solo la clase general a la cual pertenecen.

Los requerimientos funcionales que deben cumplirse para generar la distribución y evolución por temática a nivel clase de las patentes concedidas en el periodo seleccionado son los siguientes:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.

- iv. Generar el reporte de distribución general y evolución por temática a nivel clase.
- v. Mostrar los resultados

Como en el caso del reporte descrito en el epígrafe anterior, en este reporte el analista debe llenar el formulario con la opción Patentes OEMP, Comunidad, Umbral, y Periodo. Al proporcionar y aceptar estos valores se ejecuta el *script* que genera el reporte. Las relaciones involucradas en el proceso para obtener los datos solicitados son *patentoepm*, *ipcnivell*, e *ipc*.

La distribución general por temática a nivel clase se obtiene contando las patentes concedidas en el periodo seleccionado, agrupadas por temática, y se asigna un punto por temática. Para la generación de la evolución por área temática se hace un conteo similar al usado en la distribución, pero el agrupamiento se hace por área temática y año. Al igual que en los reportes anteriores, los datos son formateados y separados por el carácter (l) para mostrarse en pantalla. En la Figura 4.25 se presenta el diagrama de secuencia.

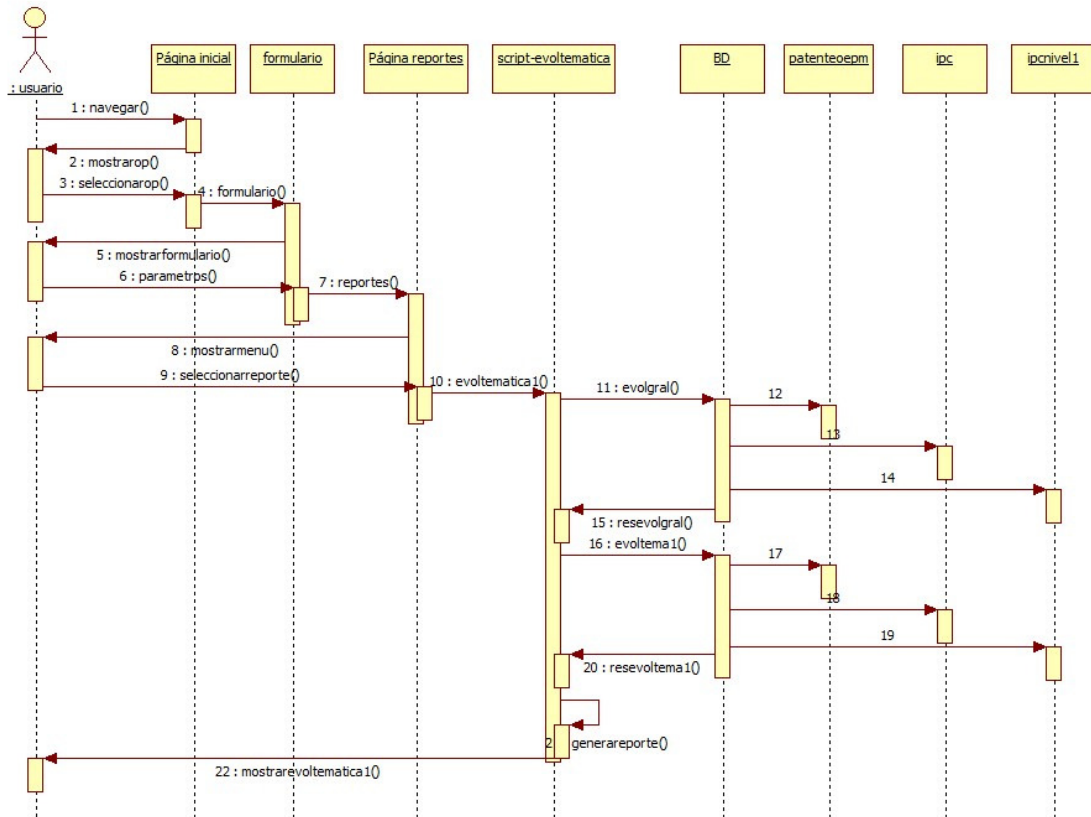


Figura 4.25: Diagrama de secuencia para generar evolución general por temática a nivel clase.

4.2.3.4. Distribución general y evolución por temática nivel subgrupo

Al igual que el reporte anterior, éste proporciona tablas de datos de la distribución por área temática y evolución de las patentes, la diferencia radica en el nivel de clasificación, aquí los datos presentados son a nivel subgrupo.

Los requerimientos funcionales que deben cumplirse para generar la distribución y evolución por temática a nivel subgrupo de las patentes concedidas en el periodo seleccionado son los siguientes:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.

- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución general y evolución por temática a nivel subgrupo.
- v. Mostrar los resultados

En la Figura 4.26 se muestra el diagrama de secuencia para este reporte, el proceso no se describe debido a que es similar al anterior, pero resulta valioso destacar que la relación utilizada en la consulta es *ipcnivel2*, la cual contiene la clasificación de las patentes a nivel subgrupo.

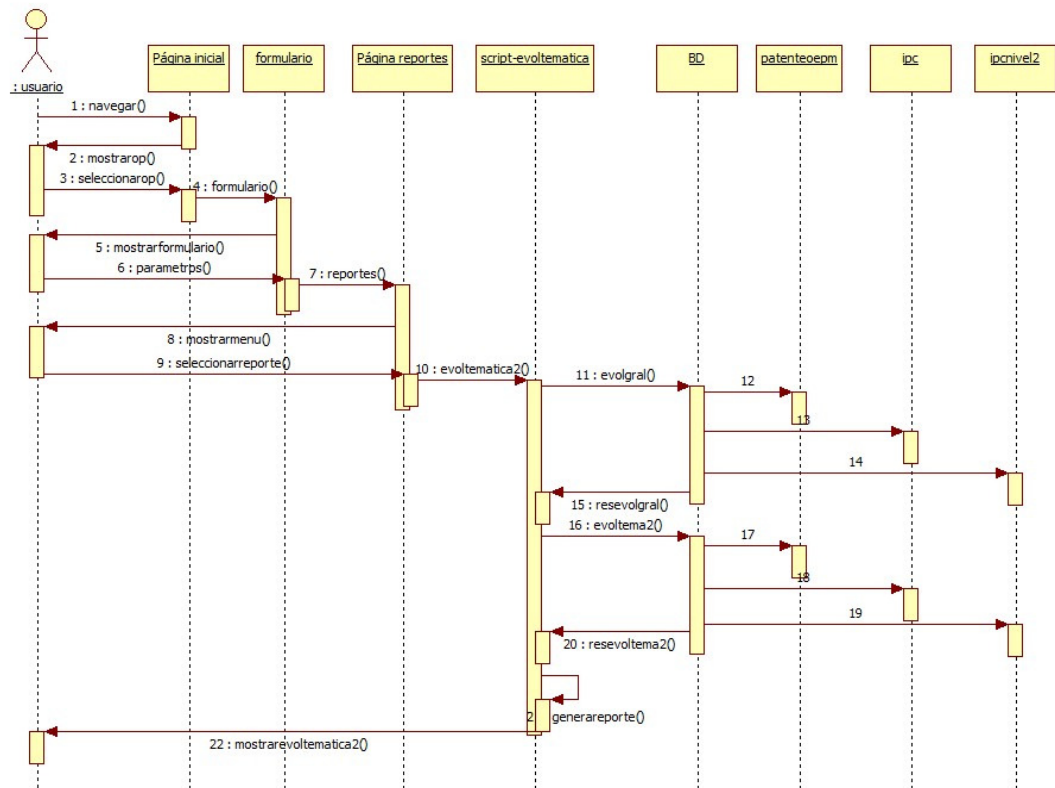


Figura 4.26: Diagrama de secuencia para generar evolución general por temática a nivel subgrupo.

4.2.3.5. Distribución por sector y temática a nivel clase y subgrupo

Los datos proporcionados en estos reportes responden a la necesidad de realizar análisis de la distribución de las patentes por sector y temática tanto a nivel clase

como a nivel subgrupo. A continuación se relacionan los requerimientos funcionales que se debe cumplir para generar este reporte.

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución general por sector y temática a nivel clase.
- v. Mostrar los resultados

Para generar este reporte el usuario llena el formulario con los datos siguientes: a) Patentes OEMP, b) Comunidad, c) Umbral, y d) Periodo, al proporcionar y aceptar estos valores es ejecutado el *script* para este propósito implementado. Las relaciones utilizadas en el proceso para obtener los datos solicitados son *patenteoepm*, *patenteinstitucion*, *tipoinstitucion*, *ipc*, *ipcnivel1* o *ipcnivel2* (dependiendo del nivel de clasificación que se obtenga).

La distribución general por sector y temática a nivel clase se calcula contando las patentes concedidas en el periodo seleccionado, se agrupan por sector y temática a nivel clase. Al obtener los datos de la consulta realizada se formatean para su presentación al usuario. En la Figura 4.27 se muestra el diagrama de secuencia de la distribución por sector y temática. El diagrama de la Figura 4.28 corresponde a la distribución por sector y temática a nivel subgrupo, en este último diagrama se observa que la relación usada para la obtención de este reporte es *ipcnivel2*.

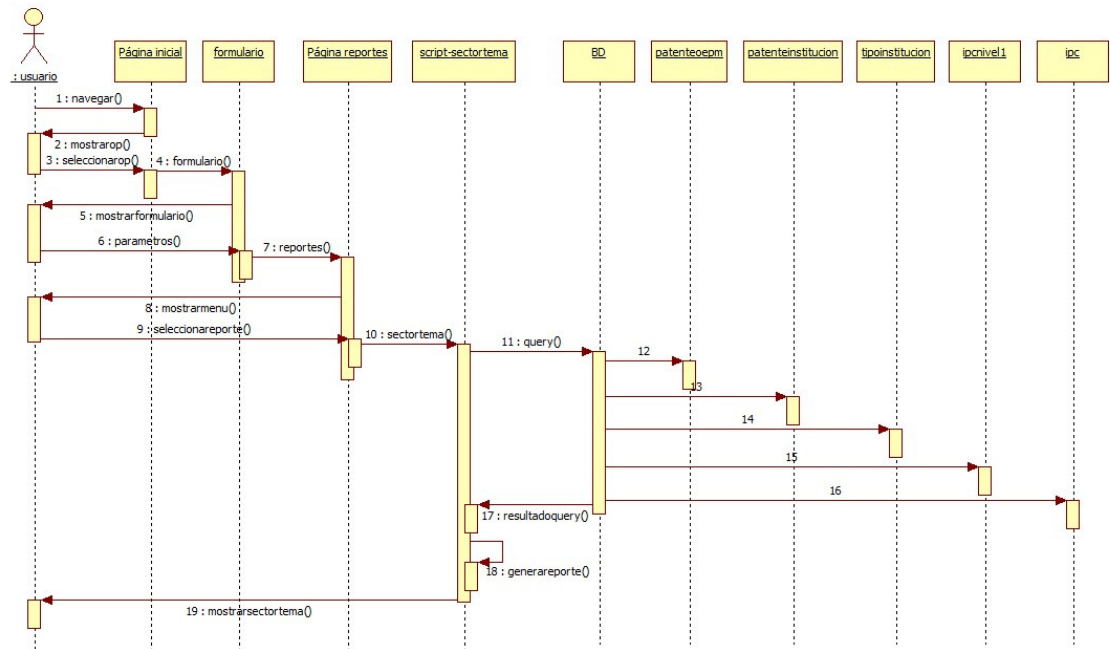


Figura 4.27: Diagrama de secuencia para generar distribución por sector y temática nivel clase.

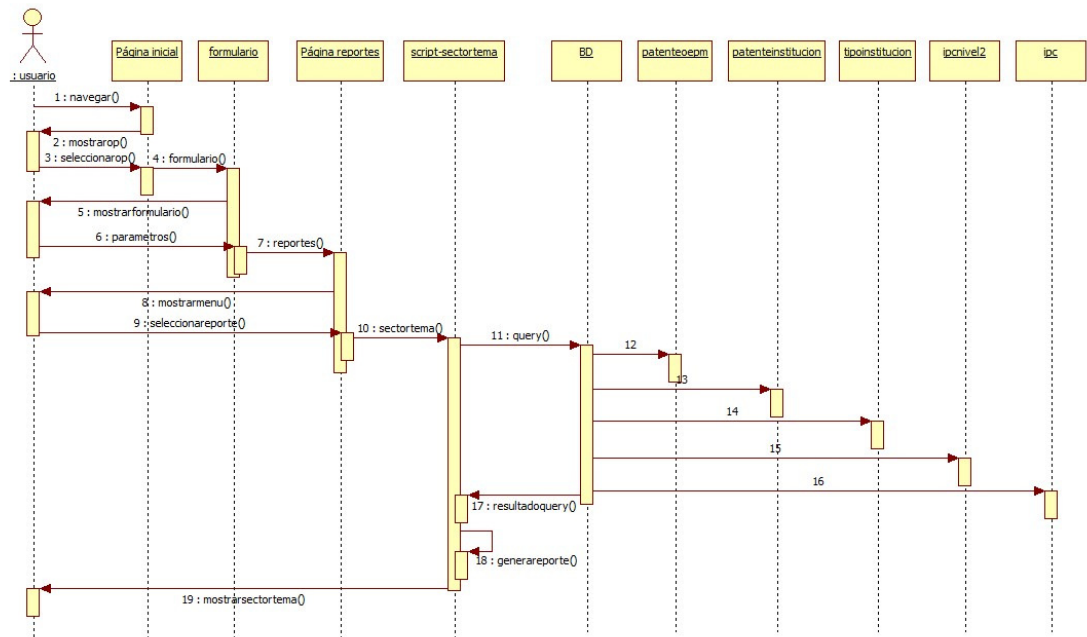


Figura 4.28: Diagrama de secuencia para generar distribución por sector y temática a nivel subgrupo.

Hasta aquí se han descritos los procedimientos para obtener reportes relacionados con patentes, en los siguientes epígrafes se describen los procedimientos para obtener reportes sobre familias de patentes.

4.2.3.6. Evolución general de las familias de patentes

La evolución general de las familias de patentes es obtenida siguiendo el mismo procedimiento que para la obtención de la evolución de las patentes concedidas, solo hay una variación en los valores proporcionados por el usuario en el formulario de parámetros de consulta. En la Figura 4.29 se muestra el diagrama de secuencia para generar este reporte.

En el formulario de parámetros de consulta mostrado en la Figura 4.21 se debe seleccionar la opción “Patentes (DERWENT)” y periodo, el resto de variables presentes en el formulario no son tomadas en cuenta en los reportes relacionados con familias de patentes. Después de completar y aceptar los valores del formulario se inicia el procesamiento de la consulta para ello se utiliza la relación *patentedwoepm*, con el fin de contabilizadas las familias de patentes concedidas dentro del periodo solicitado, se agrupan por año de concesión; los datos son mostrados formateados y mostrados en la pantalla del navegador para su posterior uso por parte del analista. En la Figura 4.29 se muestra el diagrama de secuencia para generar la evolución general de la familia de patentes.

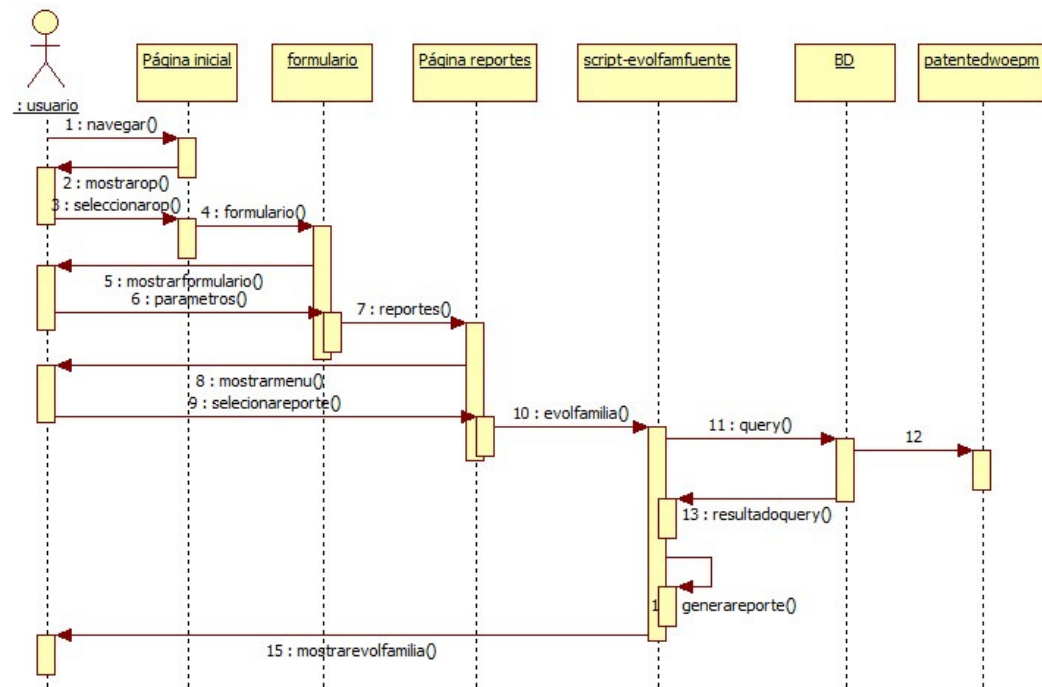


Figura 4.29: Diagrama de secuencia para generar evolución de las familias de patentes.

4.2.3.7. Evolución general de las familias de patente por tamaño

En este reporte se proporcionan datos relacionados con la distribución que tienen las familias de patentes tomando en cuenta el número de miembros que las conforman, la agrupación es anual.

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de evolución general por tamaño de las familias.
- v. Mostrar los resultados

En la Figura 4.30 se muestra el diagrama de secuencia para obtener la familia de patentes por tamaño. En el formulario de parámetros de consulta mostrado en la Figura 4.21 se debe seleccionar la opción “Patentes (*Derwent*)”, y rellenar los otros

parámetros presentes en el formulario. Después de completar y aceptar los valores solicitados en el formulario de parámetros se inicia el procesamiento de la consulta. Para ello se utiliza la relación *patentenum*, para realizar el cómputo del tamaño se utiliza el atributo *utt* ya que contiene el dato que identifica a los miembros de las familias, de tal forma que la consulta para obtener la distribución se hace contando los números de patentes con un mismo *utt*, ya agrupándolas en función de los valores contenidos en este atributo y el año de concesión (*anocon*). Los datos son mostrados formateados y mostrados en la pantalla del navegador para su posterior uso por parte del analista.

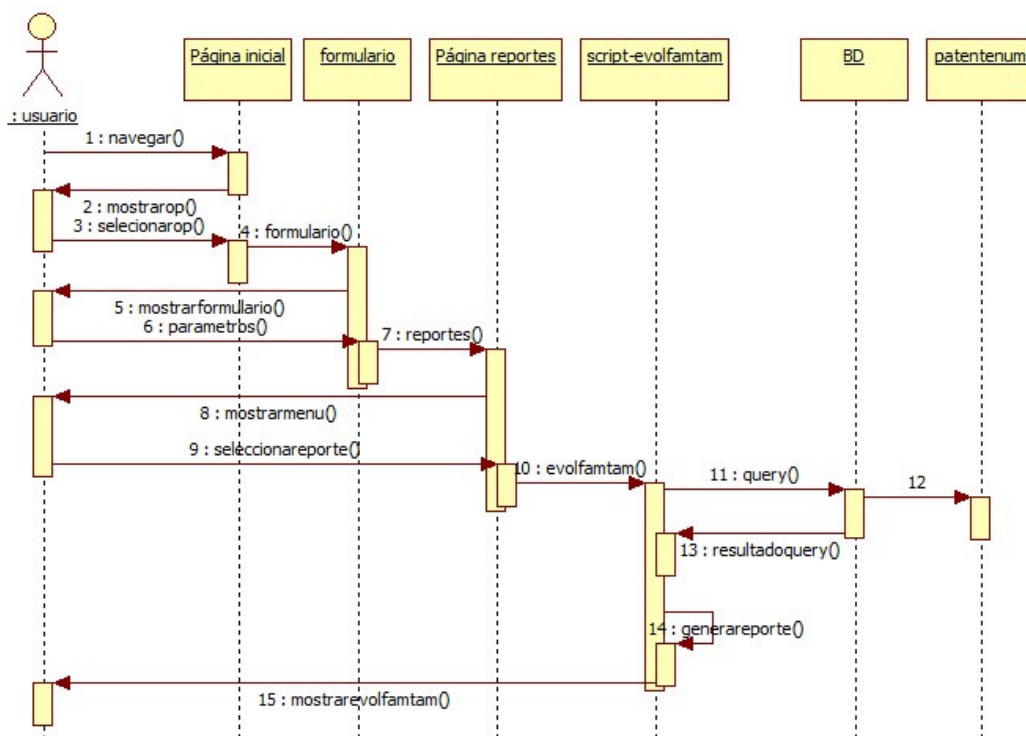


Figura 4.30: Diagrama de secuencia para generar la evolución de las familias de patentes por tamaño.

4.2.3.8. Temática de las familias de patentes a nivel clase y subgrupo

Los datos proporcionados en estos reportes responden a la necesidad de realizar análisis de la distribución de las patentes por sector y temática tanto a nivel clase

como a nivel subgrupo. A continuación se relacionan los requerimientos funcionales que se deben cumplir para generar este reporte.

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución general por temática a nivel clase.
- v. Mostrar los resultados.

Para generar reportes de este tipo el usuario selecciona la opción “Familia de patentes” y llena los otros campos del formulario con los valores requeridos, al proporcionar y aceptar estos datos se ejecuta el *script* que lleva a cabo la consulta solicitada. Las relaciones utilizadas en el proceso para obtener los datos solicitados son *patenteipc*, *ipc*, *ipcdwii1* o *ipcdwii2* (dependiendo del nivel de clasificación que se desea obtener).

La distribución de la familia de patentes por temática a nivel clase se calcula contando las patentes concedidas en el periodo seleccionado. Los datos son agrupados por temática a nivel clase utilizando la relación *ipcdwii1* y por año de concesión. Una vez que se ha realizado la consulta en el repositorio de datos se procede a dar formato a los resultados obtenidos. En la Figura 4.31 se muestra el diagrama de secuencia correspondiente a este tipo de reportes, y en la Figura 4.32 se muestra el diagrama correspondiente a la distribución por temática a nivel subgrupo, los procesos para la obtención de ambos reportes son similares. En el segundo reporte se utiliza la relación *ipcdwii2*.

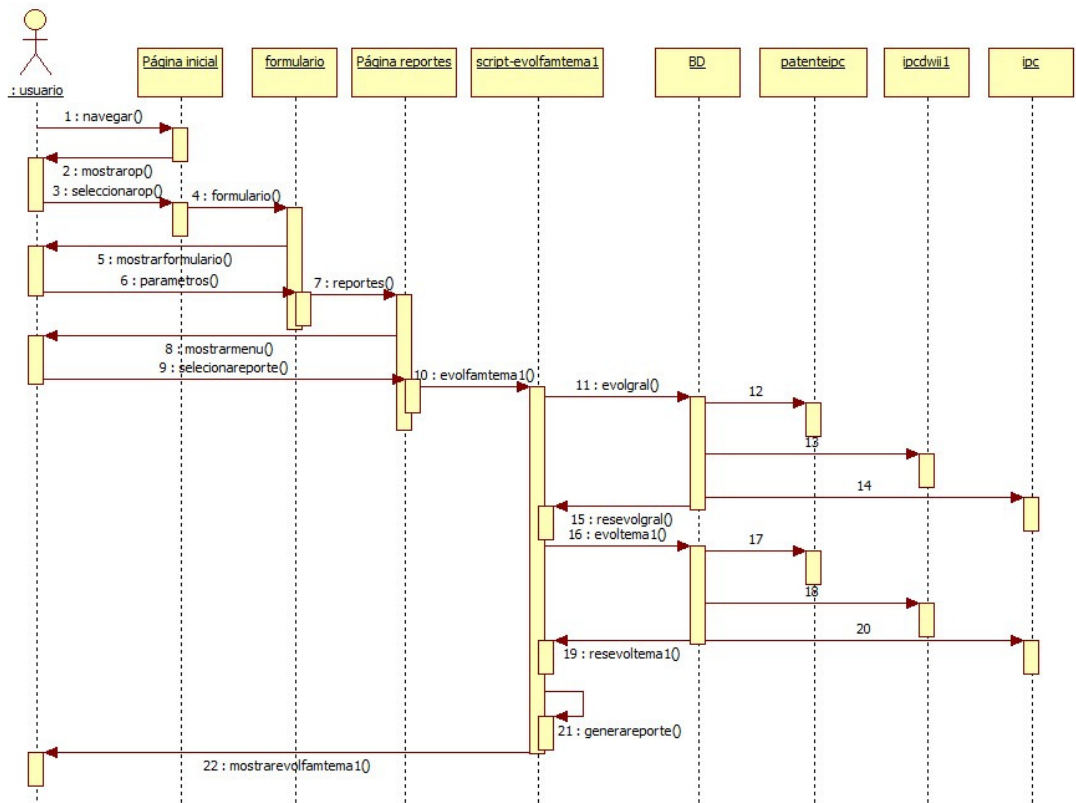


Figura 4.31: Diagrama de secuencia para generar evolución de las familias de patentes por área temática a nivel clase.

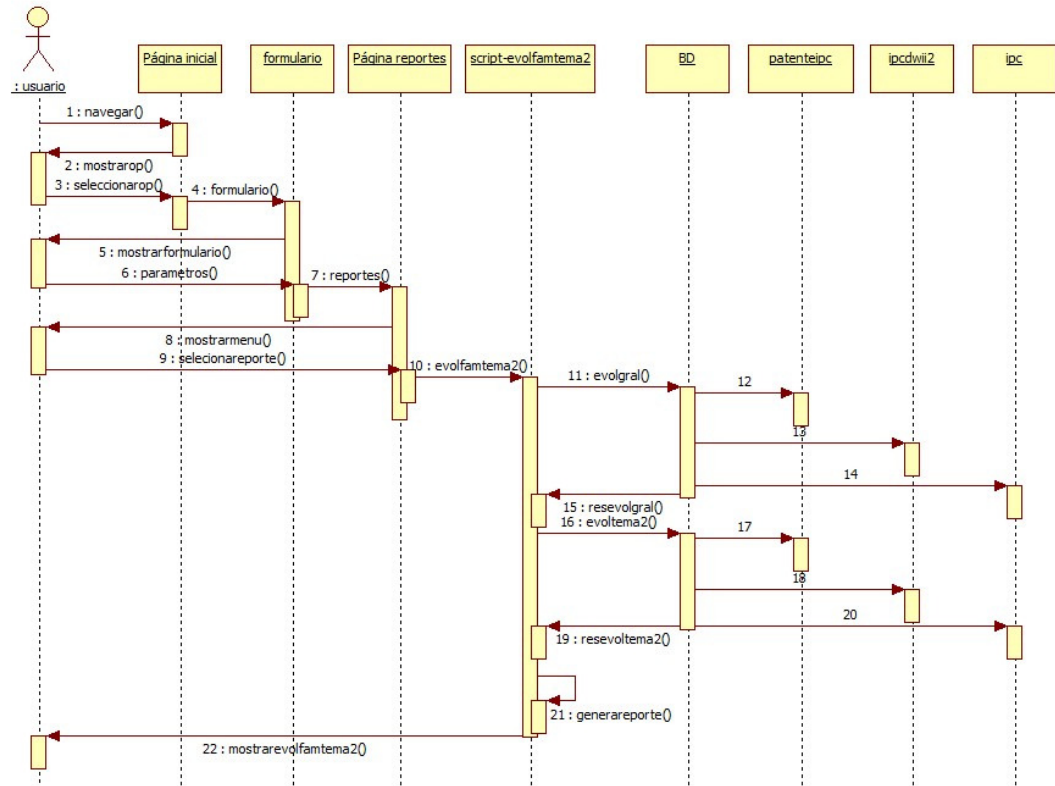


Figura 4.32: Diagrama de secuencia para generar evolución de las familias de patentes por área temática a nivel clase.

4.2.3.9. Distribución de las nacionalidades de las patentes referenciadas

En este reporte se muestra la distribución por nacionalidad de las patentes referenciadas, es importante recordar que la nacionalidad de las patentes referenciadas se obtiene de la información que proporciona la base de datos *Worldwide-espacenet* en los datos de los solicitantes de las patentes.

A continuación se relacionan los requerimientos para generar este tipo de reporte:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.

- iv. Generar el reporte de la distribución de la nacionalidad de las patentes referenciadas.
- v. Mostrar los resultados

Al igual que en los reportes anteriores relacionados con las familias de patentes, para obtener este reporte se debe seleccionar la opción de Familia de patentes en el formulario inicial. Cuando se ha llenado el formulario y aceptado los valores seleccionado, se procede a realizar el reporte solicitado haciendo la búsqueda en las relaciones *referencia* y *patenteref*. El cálculo para mostrar la distribución de la nacionalidad de las patentes referenciadas se obtiene contando y agrupando los registros por el atributo nacionalidad en el periodo solicitado. Los datos son formateados para mostrarse en pantalla (ver Figura 4.33 con el diagrama de secuencia para obtener la distribución por nacionalidad de las patentes referenciadas).

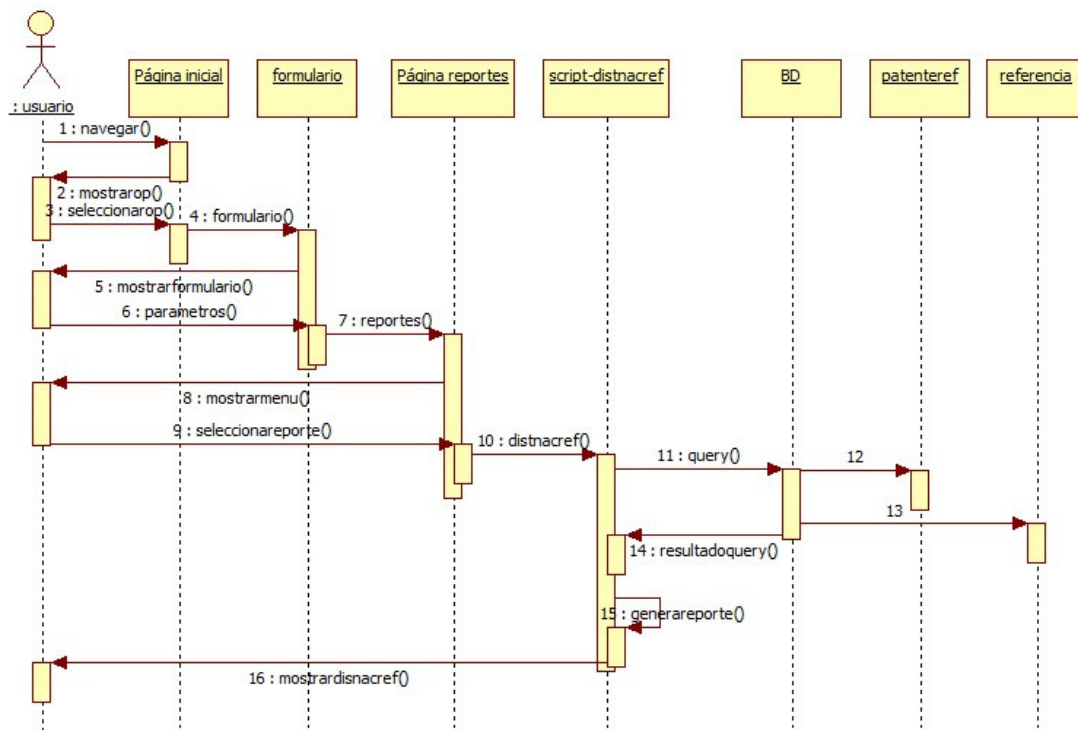


Figura 4.33: Diagrama de secuencia de la distribución por nacionalidad de las patentes referenciadas.

4.2.4.0. Temática de las patentes referenciadas a nivel clase y subgrupo

Los procedimientos para generar los reportes de las patentes referenciadas por temática a nivel clase y subgrupo son iguales a los procesos seguidos para generar los reportes de las patentes y familias de patentes, solo se debe enfatizar que las relaciones involucradas en estos procedimientos son *patenteref*, *referencia*, *patrefipc1*, *patrefipc2*, e *ipc*. En las Figuras 4.34 y 4.35 se muestran los diagramas de secuencia de distribución de la temática de las patentes referenciadas a nivel clase y subgrupo respectivamente.

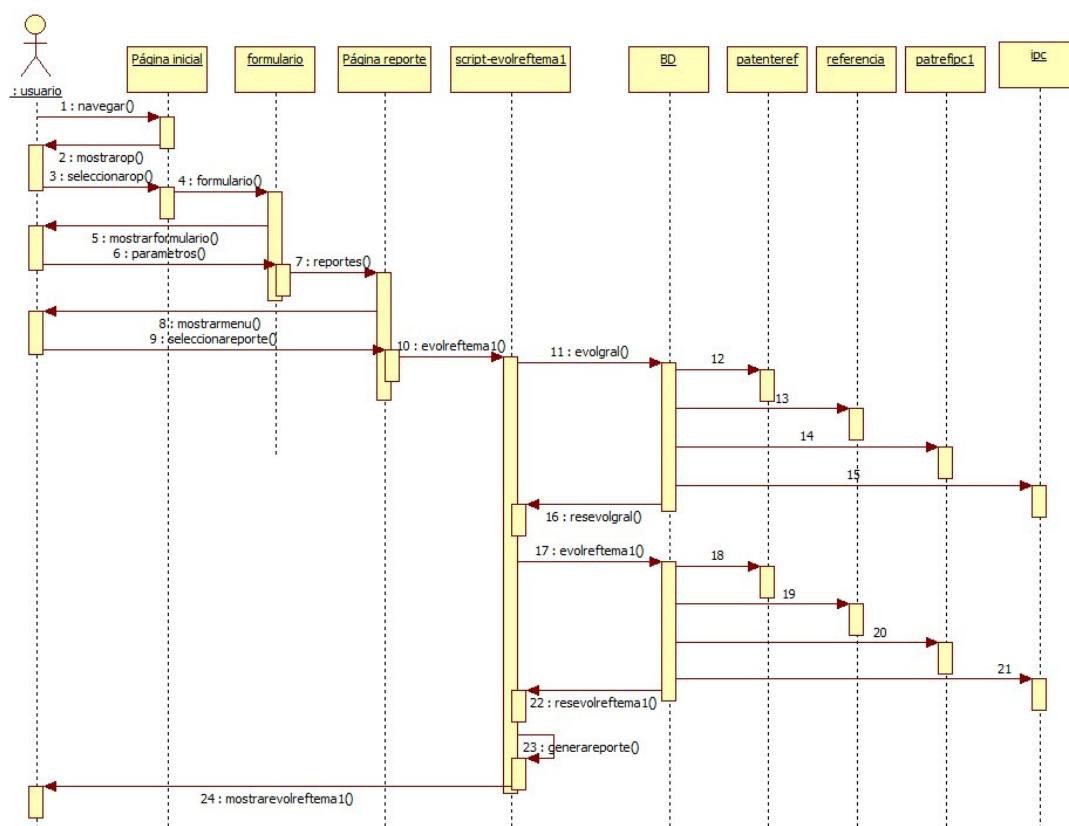


Figura 4.34: Diagrama de secuencia para generar evolución de la temática referenciada a nivel clase.

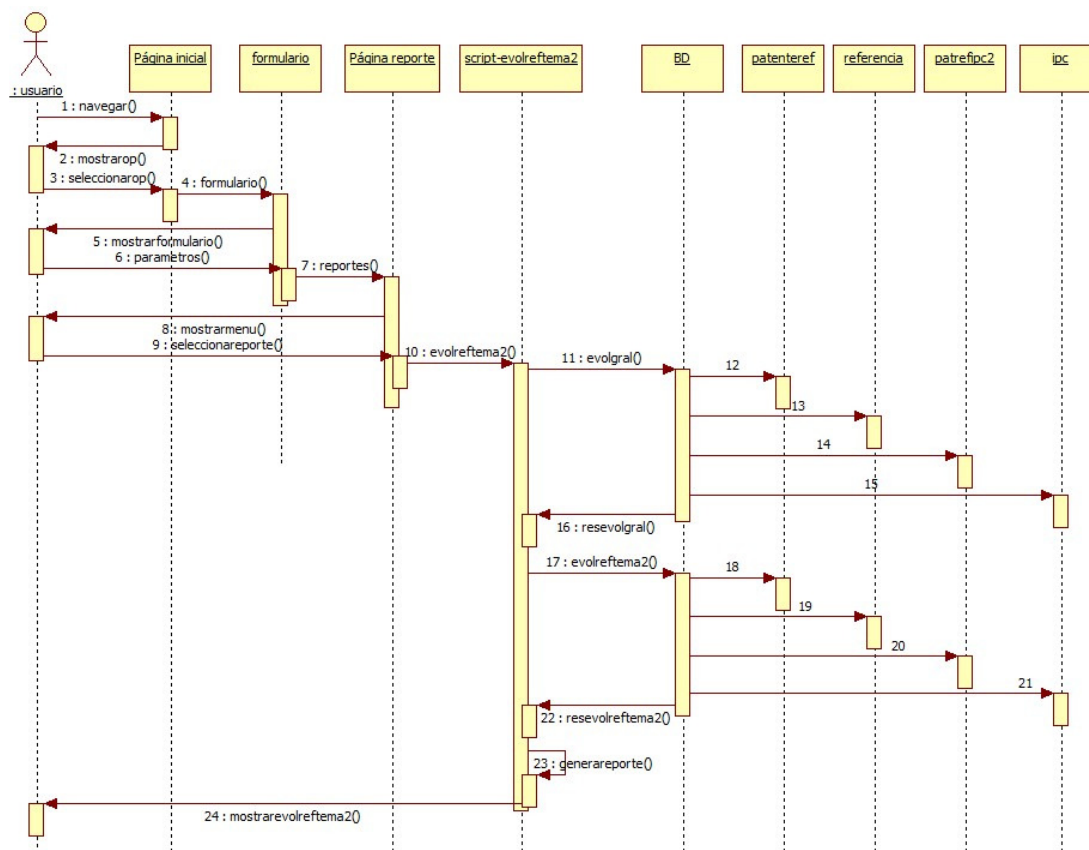


Figura 4.35: Diagrama de secuencia para generar evolución de la temática referenciada a nivel subgrupo.

4.2.4.1. Distribución de las patentes referenciadas por temática a nivel clase y nacionalidad

En este tipo de reportes se proporcionan tablas con datos que muestran la distribución de las patentes referenciadas por las familias de patentes por área temática a nivel clase y nacionalidad. A continuación se enumeran los requerimientos funcionales que se debe cumplir para generar estos reportes.

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.

- iv. Generar el reporte de distribución de patentes referenciadas por área temática y nacionalidad.
- v. Mostrar los resultados.

Estos reportes también pertenecen a los relacionados con las familias de patentes debido a que son patentes referenciadas por las familias, así, para obtenerlos el usuario selecciona la opción “Patentes (*Derwent*)” y llena los otros campos del formulario con los valores requeridos (mostrados en la página de parámetros de la Figura 4.21). En el momento de aceptar estos valores se ejecuta el *script* implementado para realizar la consulta. Las relaciones utilizadas en el proceso para obtener los datos solicitados son *patenteref*, *referencia*, *ipcpatref1*, e *ipc*.

La distribución de las patentes referenciadas por temática a nivel clase y nacionalidad se obtiene contando las patentes referenciadas en el periodo seleccionado. Los datos son agrupados por año de concesión y temática a nivel clase utilizando la relación *ipcpatref1*. Una vez que se ha realizado la consulta en el repositorio de datos se procede a dar formato a los resultados obtenidos y se muestran al usuario (ver diagrama de la Figura 4.36).

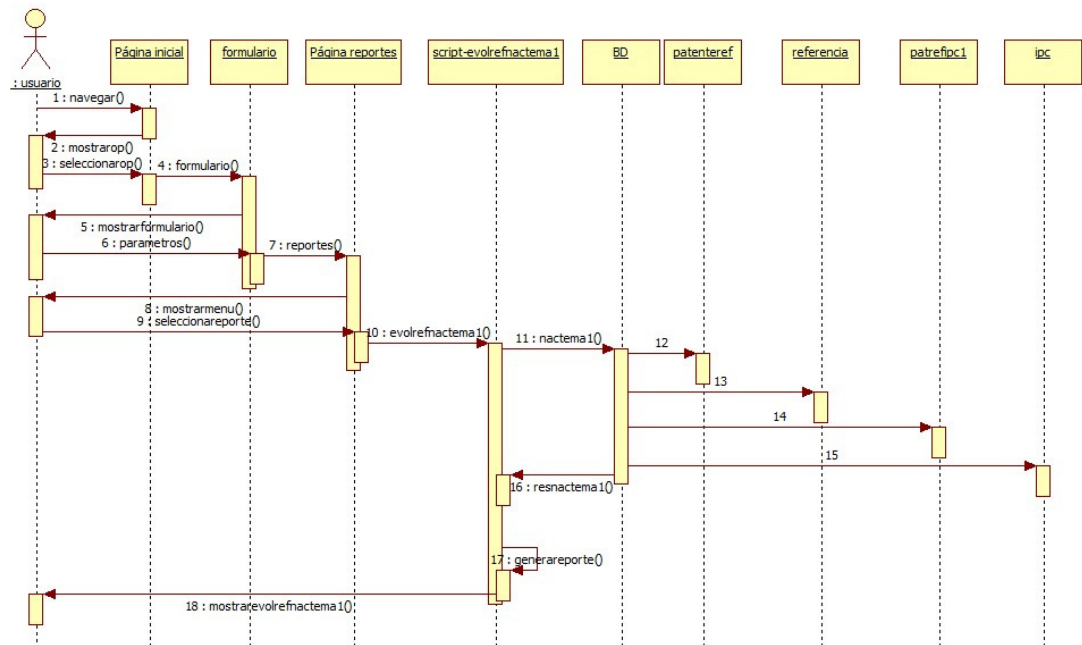


Figura 4.36: Diagrama de secuencia para generar distribución de las patentes referenciadas por temática y nacionalidad.

4.2.4.2. Distribución de la temática de las patentes fuente y temática de las patentes referenciadas

Este tipo de reportes se utiliza para mostrar la relación existente entre la temática de las familias de patentes citantes y la temática de las patentes referenciadas. A los resultados obtenidos en este tipo de consultas se les da un tipo de formato aceptado por las herramientas de visualización *Gephi*, *Pajek*, y *VOSviewer*.

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución por áreas temáticas de las patentes fuente y patentes referenciadas.
- v. Mostrar los resultados

Las relaciones utilizadas en el proceso para obtener los datos solicitados en los reportes de distribución de las temáticas de patentes fuente y patentes referenciadas a nivel clase son: *patenteipc*, *ipcdwii1*, *ipc*, *referencia*, e *ipcpatref1*.

La distribución de las temáticas a nivel clase de las patentes fuente y las patentes referenciadas se obtiene contando para cada una de las temáticas de las patentes fuente las temáticas de las patentes referenciadas. Los datos son agrupados por año de concesión, temática a nivel clase de las patentes fuente (*ipcdwii1*), y temática de las patentes referenciadas (*ipcpatref1*). Una vez que se ha realizado la consulta en el repositorio de datos se procede a dar formato a los resultados obtenidos y se muestran al usuario. La Figura 4.37 muestra el diagrama de secuencia para este reporte.

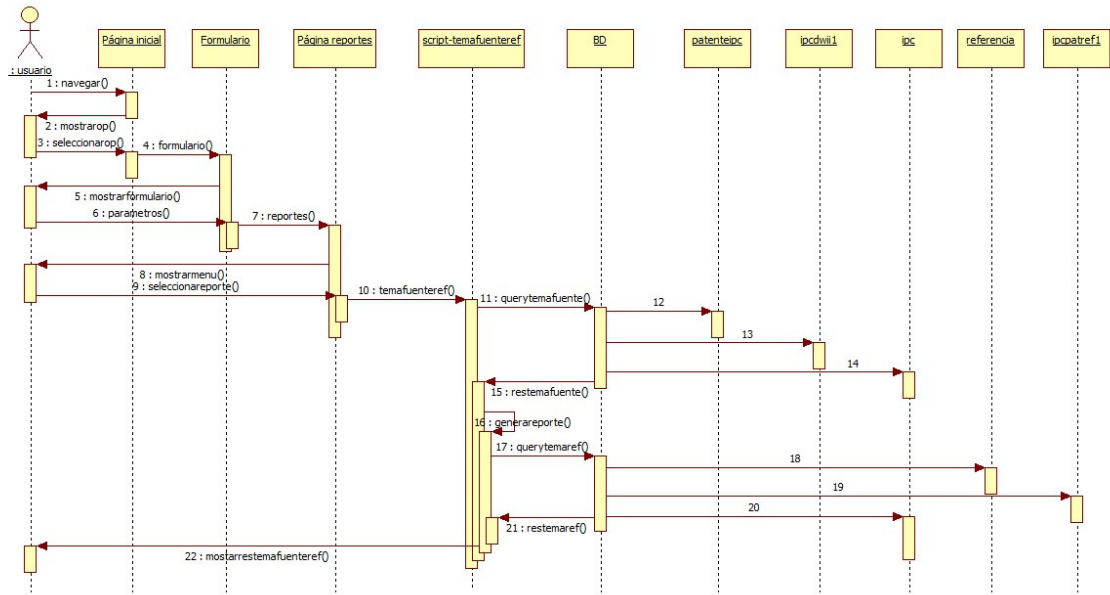


Figura 4.37: Diagrama de secuencia temática fuente y temática referenciada.

4.2.4.3. Evolución de las familias de patentes triádicas

La evolución de las familias de patentes triádicas es similar a los reportes de la evolución de las familias de patentes, pero en este caso es un reporte sobre un tipo de familias específicas ya que solo se contabilizan aquellas familias que dentro de sus miembros se deben contar con números de patentes de las USPTO, EPO, y JPO.

Los requerimientos funcionales para obtener estos reportes son:

- i. Mostrar las opciones disponibles por el módulo.
- ii. Mostrar un formulario solicitando parámetros generales para realizar la búsqueda.
- iii. Mostrar una lista de reportes disponibles.
- iv. Generar el reporte de distribución de las familias de patentes triádicas.
- v. Mostrar los resultados

Al pertenecer este reporte al grupo de las familias de patentes, en el formulario de parámetros de consulta de la Figura 4.21 se selecciona la opción “Patentes (*Derwent*)” y periodo, en este caso no se toma en cuenta el umbral debido a que el

número de familias triádicas es pequeño. Al completar y aceptar los valores del formulario se inicia el procesamiento de la consulta, para obtener esta distribución se utiliza la relación *triadicas*, que permiten contabilizar las familias de patentes triádicas concedidas dentro del periodo solicitado, se agrupan por año de concesión. Los datos son mostrados formateados y mostrados en la pantalla del navegador para su posterior uso por parte del analista (ver Figura 4.38).

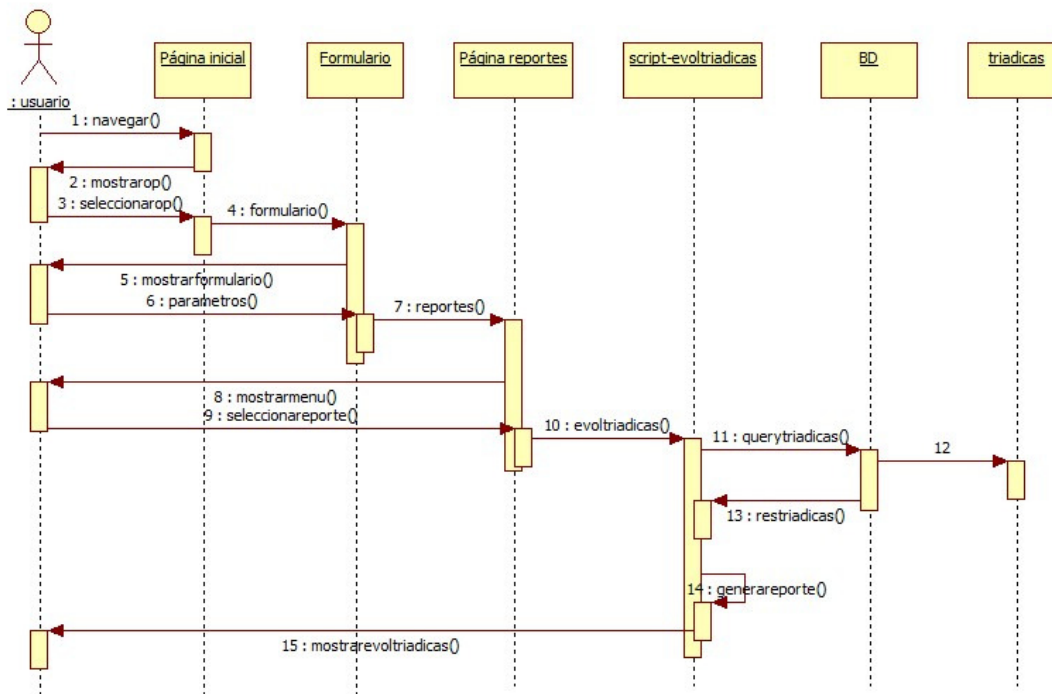


Figura 4.38: Diagrama de secuencia de evolución de familias triádicas.

4.2.4.4. Distribución de las familias de patentes triádicas por área temática

Los reportes para obtener la distribución de las familias de patentes triádicas por área temática siguen la misma lógica utilizada en los procesos para obtener este tipo de reportes en las patentes, familias de patentes y patentes referenciadas. En este caso en particular cabe destacar que *triadicas*, *ipcdwii1*, e *ipc* son las relaciones utilizadas en el procesamiento de los datos. En la Figuras 4.39 se muestra el diagrama de secuencia de distribución de la temática de las familias de patentes triádicas a nivel clase.

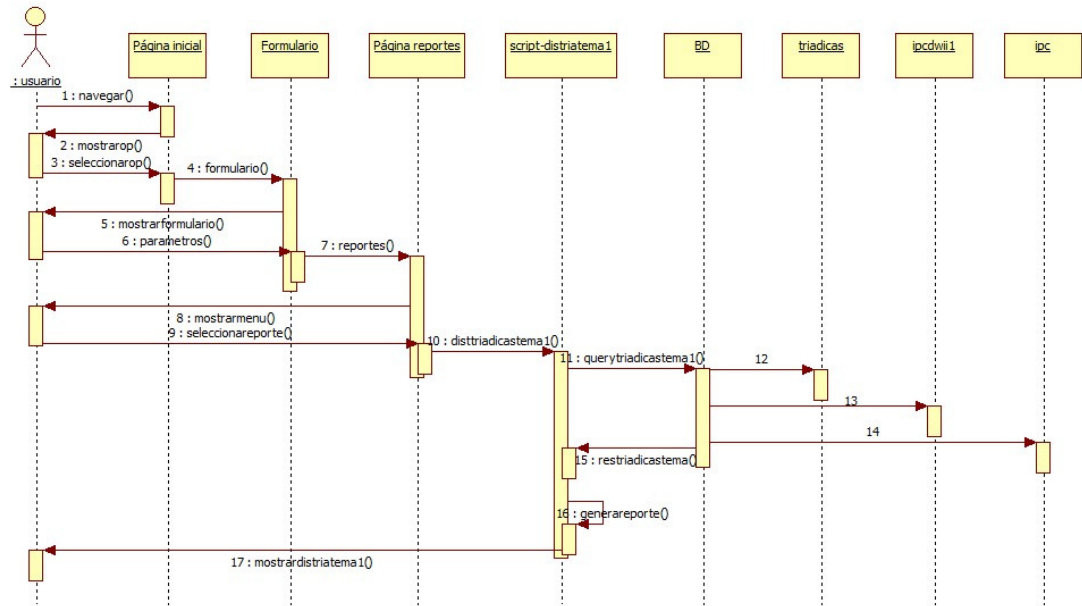


Figura 4.39: Diagrama de secuencia de familias triádicas por área temática.

Los tipos de reportes hasta aquí mostrados, así como los *scripts* para el tratamiento de datos se probaron en un conjunto de datos de patentes de españolas concedidas en el periodo 1997-2008. Los resultados obtenidos en este estudio de caso se describen en la sección siguiente. También es importante mencionar que con el modelo de datos diseñado es posible obtener más indicadores de los descritos anteriormente.

4.3. Aplicación del Modelo teórico-metodológico: Estudio de caso con una muestra de patentes españolas

El modelo teórico-metodológico presentado en esta investigación ha sido probado en un conjunto de patentes españolas concedidas en el periodo 1996 a 2008 (22.215 patentes), los resultados obtenidos utilizando esta propuesta teórico-metodológica tiene como propósito mostrar la obtención de indicadores a partir de los reportes proporcionados en la metodología y no la de hacer la caracterización de las patentes de españolas. También es importante mencionar que los registros de estas patentes fueron conseguidos a través de la OEPM, ya que la interfaz de búsqueda proporcionada por INVENES.

4.3.1 Capacidad innovadora de la Comunidad de Madrid a partir de las patentes concedidas entre 1996 y 2007

El propósito de este apartado es, además de probar el modelo teórico-metodológico, conocer algunas de las adecuaciones en el procesamiento de los datos sobre las características de la actividad patentadora de la Comunidad de Madrid (CM) con respecto al conjunto de España. Para ello, se analizaron las patentes concedidas entre 1996 y 2007 recogidas en la base de datos *Interpat* de la Oficina Española de Patentes y Marcas.

En un trabajo titulado “*Capacidad innovadora de la Comunidad de Madrid a partir de las patentes concedidas entre 1996 y 2007*” (Lascurain et al, 2010) se involucró la utilización de otro tipo de indicadores como son los recursos humanos y materiales de la CM. Sin embargo, como ya se mencionó anteriormente, aquí se persigue mostrar la utilización de los indicadores obtenidos a partir de los reportes generados con la metodología propuesta, es por este motivo que únicamente se muestra en este apartado los indicadores relativos a las patentes en el periodo 1996-2007 (20.433 patentes).

Las variables consideradas fueron: año de concesión, nombre de los solicitantes, clasificación internacional de patentes (a nivel tanto de secciones como de clases), sectores institucionales y provincia del solicitante para delimitar el análisis a la CM.

Para la elaboración de tablas de contingencia y a la obtención de indicadores unidimensionales relativos a producción tecnológica, evolución anual de la producción, sectores institucionales y temáticas, así como a los indicadores multidimensionales se utilizó la sección de reportes utilizando la opción “Patentes OEPM” del formulario de parámetros mostrado en la Figura 4.21. En este caso, se empleó el Análisis de Correspondencias (AC) para relacionar las fechas de concesión y los sectores, y estos últimos con las temáticas. Este análisis fue realizado con el programa estadístico Xlstat.

4.3.1.1 Las invenciones en la CM y en el conjunto de España

Durante el período analizado, el número total de patentes concedidas en España ha tenido un incremento medio anual del 17,22 %, mientras que el observado en la CM fue inferior, concretamente del 11,72 % (Tabla 4.6). Estos datos son llamativos ya que muestran una menor actividad patentadora por parte de la CM con respecto al conjunto español en los últimos años

El análisis del número de patentes concedidas en España y en la CM a lo largo del período de estudio muestra que la región madrileña ha sido responsable del 20,74 % de las concedidas en el conjunto del país. En este sentido, es importante ver la evolución de la contribución de la CM al conjunto español, puesto que esta comienza representando el 25,32 % de total en 1996, para descender al 16,30 % en 2007. Evidentemente esto podría deberse al fuerte incremento de las patentes concedidas al resto de comunidades autónomas, que hace que la CM pierda peso en el conjunto nacional.

Tabla 4.6: Número de patentes concedidas en España y en la CM

Año	Patentes España	Incremento anual	Patentes CM	Incremento anual	CM respecto a España (%)
1996	778		197		25.32
1997	1201	54.37	284	44.16	23.65
1998	1762	46.71	399	40.49	22.64
1999	1838	4.31	412	3.26	22.42
2000	1610	-12.40	372	-9.71	23.11
2001	1774	10.19	382	2.69	21.53

2002	921	-48.08	190	-50.26	20.63
2003	1803	95.77	337	77.37	18.69
2004	1860	3.16	366	8.61	19.68
2005	2380	27.96	412	12.57	17.31
2006	2028	-14.79	358	-13.11	17.65
2007	2478	22.19	404	12.85	16.30
Total	20433		4113		
Promedio	1702.75	17.22	342.75	11.72	20.74

Fuente: OEPM. Elaboración propia.

4.3.1.2 Características de la invención en la CM

La Tabla 4.7 muestra la evolución de las patentes en la CM según los sectores institucionales contemplados en este análisis. Dicha asignación por sectores tiene como limitación el que los derechos de las patentes surgidos en el seno de una determinada institución (por ejemplo en una universidad) pueden ser solicitados por personas físicas o jurídicas ajenas a ésta.

Tabla 4.7: Evolución de patentes en la CM según sectores institucionales e incrementos anuales (entre paréntesis)

Año	Empresas	Particulares	CSIC	Universidades	OPI	Otros
1996	103	60	20	9	5	4
1997	156 (51,46)	82 (3,67)	25 (25,00)	20 (122,22)	6 (20,00)	3 (-25,00)
1998	181 (16,03)	155 (8,90)	24 (-4,00)	34 (70,00)	5 (-16,67)	10 (233,33)
1999	181 (0)	148 (-0,45)	56 (133,33)	32 (-5,88)	6 (20,00)	11 (10,00)
2000	179 (-1,10)	114 (-2,30)	58 (3,57)	25 (-21,88)	5 (-16,67)	7 (-36,36)
2001	168 (-6,15)	136 (1,93)	36 (-37,93)	54 (116,00)	6 (20,00)	5 (-28,57)
2002	65 (-61,31)	62 (-5,44)	41 (13,89)	36 (-33,33)	2 (-66,67)	4 (-20,00)
2003	141 116,92	105 6,94	55 34,15	47 30,56	8 300,00	4 0,00
2004	158 (12,06)	97 (-0,76)	78 (41,82)	48 (2,13)	9 (12,50)	7 (75,00)
2005	164 (3,80)	136 (4,02)	90 (15,38)	41 (-14,58)	10 (11,11)	9 (28,57)
2006	144 (-12,20)	101 (-2,57)	62 (-31,11)	63 (53,66)	11 (10,00)	17 (88,89)
2007	157 (9,03)	127 (2,57)	71 (14,52)	58 (-7,94)	14 (27,27)	12 (-29,41)
Total	1797 (11,68)*	1323 (1,50)*	616 (18,97)*	467 (28,27)*	87 (29,17)*	93 (26,95)*

* Promedios anuales. Fuente: OEPM. Elaboración propia.

Como era de esperar, las empresas y los particulares son los sectores que acumulan un mayor número de patentes a lo largo de los 12 años de estudio.

En valores porcentuales ambos sectores suman el 71,18 % del total, mientras que las instituciones más ligadas a la investigación, como son las universidades, CSIC y otros OPI acumulan el 26,69 %. Sin embargo, hay que destacar el gran incremento medio anual de las patentes concedidas a estos sectores a lo largo del período estudiado. Los otros OPI, presentan un incremento medio anual del 29,17 %, las universidades del 28,27 % y CSIC del 18,97 %. Estos importantes aumentos, aunque partiendo de valores bajos, vendrían a evidenciar el interés cada vez mayor que tienen los sectores mencionados por la generación de este tipo de conocimiento. Con respecto al CSIC, ha de tenerse en cuenta que el total de las patentes se gestiona desde la organización central de Madrid, aunque pueda proceder de institutos de otras Comunidades Autónomas.

En cuanto al sector empresarial, el promedio de variación anual se incrementa en el 11,68 %. Este valor se encuentra en consonancia con el incremento promedio anual de gasto de I+D ejecutado por este sector en la CM que ha sido del 12,04 % en el período considerado. También es preciso señalar el cambio de tendencia que se observa entre el sector empresarial y las universidades, el CSIC o los otros OPI, que han incrementado notablemente su actividad patentadora. Una posible explicación a esta situación podría darse por el importante aumento de la cooperación que se está produciendo entre las empresas y los centros públicos de investigación, sobre todo al aumentar la complejidad tecnológica de la innovación, tal y como se reconoce en el trabajo de Fernández de Lucio y otros (1983).

A partir de los números índice, en los que se ha tomado como base el primer año de estudio (1996), se ha podido comprobar cómo las universidades, pese a su escasa presencia en valores absolutos al inicio del período, constituyen el sector que presenta un mayor crecimiento, que alcanza en el año 2006 el 722,22 %. El CSIC, sector también vinculado a la investigación, experimenta asimismo crecimientos importantes con respecto al año base, como el del 450 % para el año 2005. En cuanto a otros OPI, los valores obtenidos oscilan entre un crecimiento del 280 % en el último año del estudio al decremento del 40 % en 2002 y el estacionamiento de los años 1998 y 2000.

Las empresas presentan un crecimiento que no supera el 175 % (años 1998 y 1999) y un año, el 2002, con un decremento del 63,11 %. Las patentes concedidas anualmente a los particulares crecen a lo largo del período de estudio, con un máximo del 258,33 % en el año 1998. Los datos se muestran en la Figura 4.40.

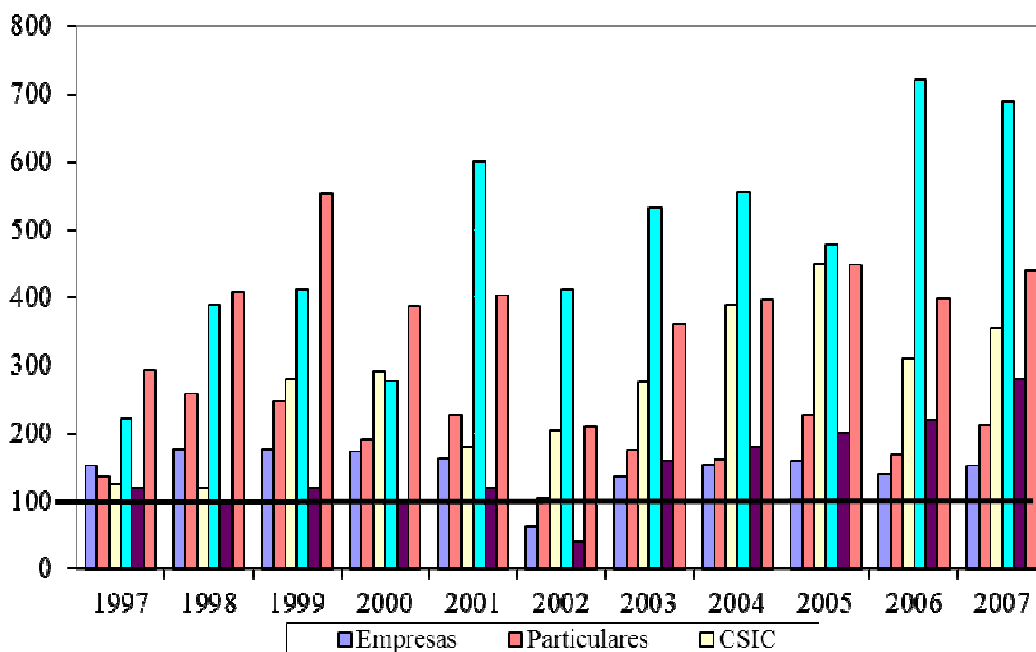


Figura 4.40: Incremento del número de patentes concedidas por sectores (números índice. Año base 1996)

En la Figura 4.41 se muestra la evolución de la actividad patentadora de los distintos sectores institucionales de la Comunidad de Madrid a lo largo de los años descritos. Para ello se ha realizado un Análisis de Correspondencias (AC) que se ha representado mediante un gráfico de burbujas. El tamaño de las burbujas en el caso de la variable «sector» es indicativo de la proporción de patentes concedidas en las categorías de la misma. La posición en el mapa informa sobre los perfiles evolutivos que muestra cada uno de los sectores. En este sentido, tanto empresas como particulares que se sitúan próximos a la zona central del mapa presentan un perfil similar a lo largo de los años estudiados, mientras que las universidades, en el cuadrante inferior derecho del mapa junto con «otros», concentra su actividad en los últimos años del estudio (2006 y 2007). Por otro lado, el CSIC y otros OPI (cuadrante superior derecho) tienen una mayor actividad en los años 2004 y 2005.

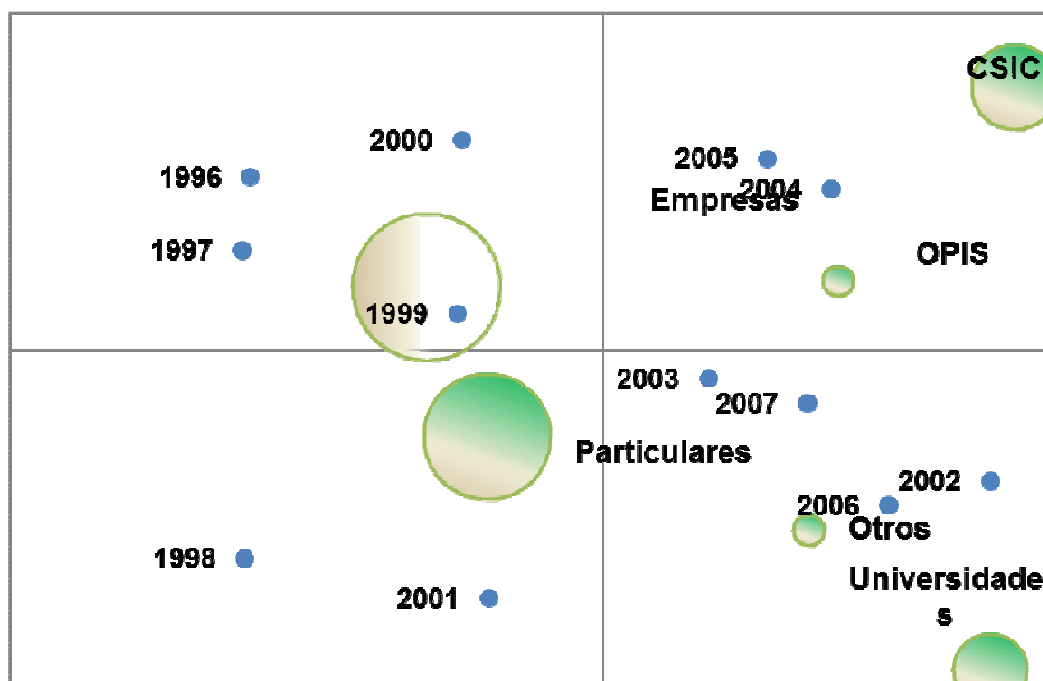


Figura 4.41: Evolución de la actividad patentadora de los sectores (análisis de correspondencias)

Otro de los aspectos que se ha querido conocer en este trabajo es la temática en la que se patenta en la Comunidad de Madrid. En este sentido, y para determinar las categorías de esta variable, se ha utilizado la Clasificación Internacional de Patentes (CIP), incluyéndose todos los códigos de clasificación asignados a cada patente.

En la Tabla 4.8 se muestran los datos correspondientes a la asignación de las ocho secciones del CIP. La sección A, «Necesidades de la Vida» que incluye las subsecciones de Actividades rurales; Alimentación y Tabaco; Objetos personales o domésticos; Salud, Protección y Farmacia, recoge el 23,56 % de las patentes. Le sigue, con un porcentaje similar (el 23,36 %) la sección C, «Química/Metal». Sólo estas dos secciones abarcan casi la mitad de las patentes concedidas en todo el período considerado (el 46,92 %).

Tabla 4.8: Temática de las patentes concedidas a la CM

Secciones	Frecuencia	%
A: Necesidades de la vida	1.022	23,56
C: Química/Metal	1.013	23,36
G: Física	969	22,34
B: Técnicas industria/Transportes	956	22,04
H: Electricidad	603	13,90
E: Construcción	406	9,36
F: Mecánica	346	7,98
D: Textiles/Papel	44	1,01
Total	4,337	

Fuente: OEPM. Elaboración propia.

La clasificación de «Física» (G), que incluye Instrumentos y Ciencia Nuclear, ocupa el tercer lugar con un porcentaje del 22,34 %.

«Técnicas industriales y transportes» (B), clasificación que abarca los temas de Separación y Mezcla, Conformación Mecánica; Imprenta; Transportes; Almacenamiento; Tecnología de las microestructuras y Nanotecnología, ocupa el cuarto lugar con el 22,04 %. «Electricidad» (H) representa el 13,90 %. Las siguientes temáticas no llegan al 10 % de las invenciones cada una de ellas, como son: «Construcción» (E) (Edificios y perforación del suelo y explotación minera) que ocupa el sexto lugar con el 9,36 %, «Mecánica» (F) con el 7,98 %, y la temática «Textiles y Papel» (D) se limita a una aportación del 1,01 % del total.

La evolución de las secciones a lo largo del período analizado se muestra en la Figura 4.42, y ha sido obtenida mediante un análisis de correspondencias. Los temas tratados que tienen una frecuencia similar en todos los años analizados son: Necesidades de la Vida, Técnicas industria/Transportes, Construcción, Química/ Metalurgia y Física. Respecto a las otras temáticas, es interesante señalar los casos de Textiles/Papel y Electricidad, que se sitúan en el cuadrante superior e inferior derecho del mapa, debido a que su actividad patentadora fue más activa a finales de la década de los noventa.

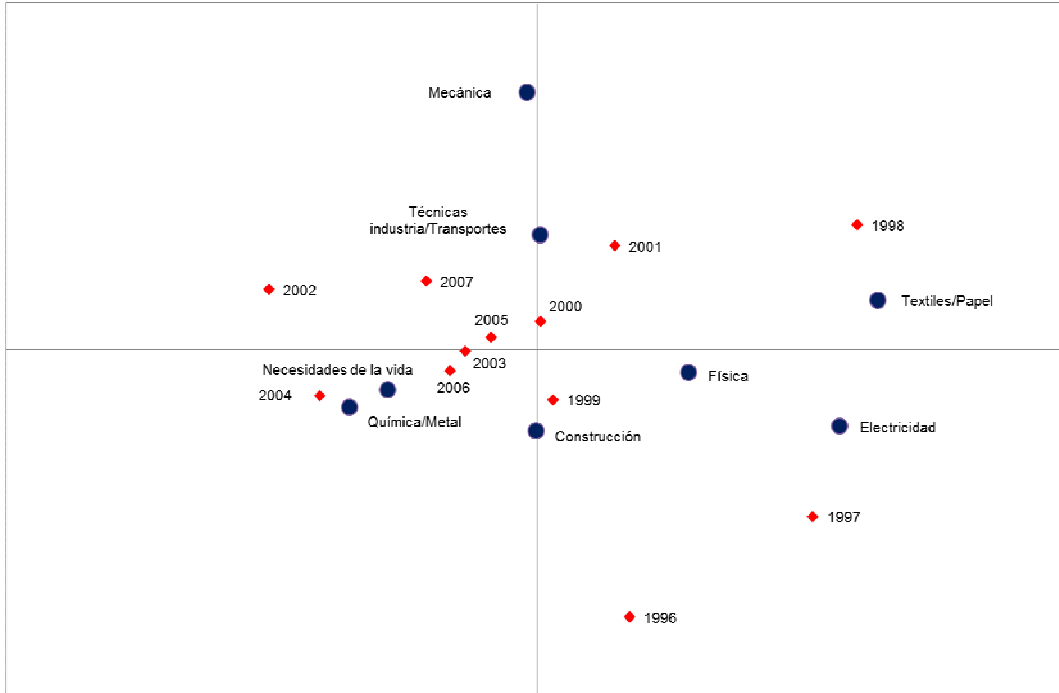


Figura 4.42: Evolución temporal de los temas en los que se patenta en la CM (Análisis de correspondencias)

Con el fin de establecer la relación entre temas en que se patenta y los distintos sectores de la CM, se ha realizado un análisis de correspondencias. La representación gráfica obtenida (Figura 4.43) permite observar la diferenciación en la forma de agruparse los sectores institucionales en torno a las temáticas sobre las que versan sus patentes. Las instituciones con mayor actividad investigadora (universidades, CSIC y otros OPI) tienen un perfil más específico que el resto, y que gira en torno a los temas agrupados en la clasificación de Química/Metalurgia. Mientras que los sectores que más patentan (empresas y particulares) se sitúan próximos al resto de las temáticas, lo que indicaría un perfil más difuso.

En la Figura 4.43 llama la atención el caso de la Física que aparece en el mapa en la posición central muy vinculada al mundo empresarial, pero en el mismo cuadrante y también cercana a las universidades, debido a la importante actividad de estas instituciones académicas en relación con los temas de esta sección.

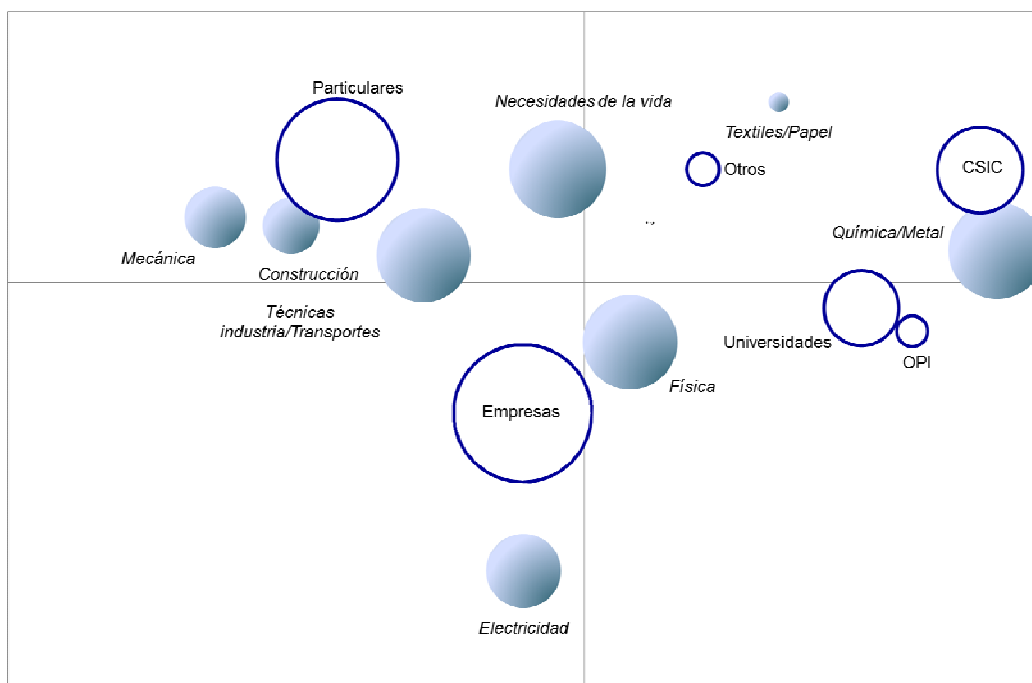


Figura 4.43: Análisis de correspondencias. Temas (a nivel sección) vs sectores

Una información más detallada sobre los temas de mayor interés a la hora de patentar se obtiene a partir del análisis de las clases de la clasificación CIP con más patentes concedidas.

Con el fin de facilitar la visualización de los datos sólo se han considerado aquellos temas en los que aparecía un número de patentes igual o superior a 30 (Figura 4.44). En este mapa no figuran los otros OPI ya que no cumplen el requisito establecido en el umbral.

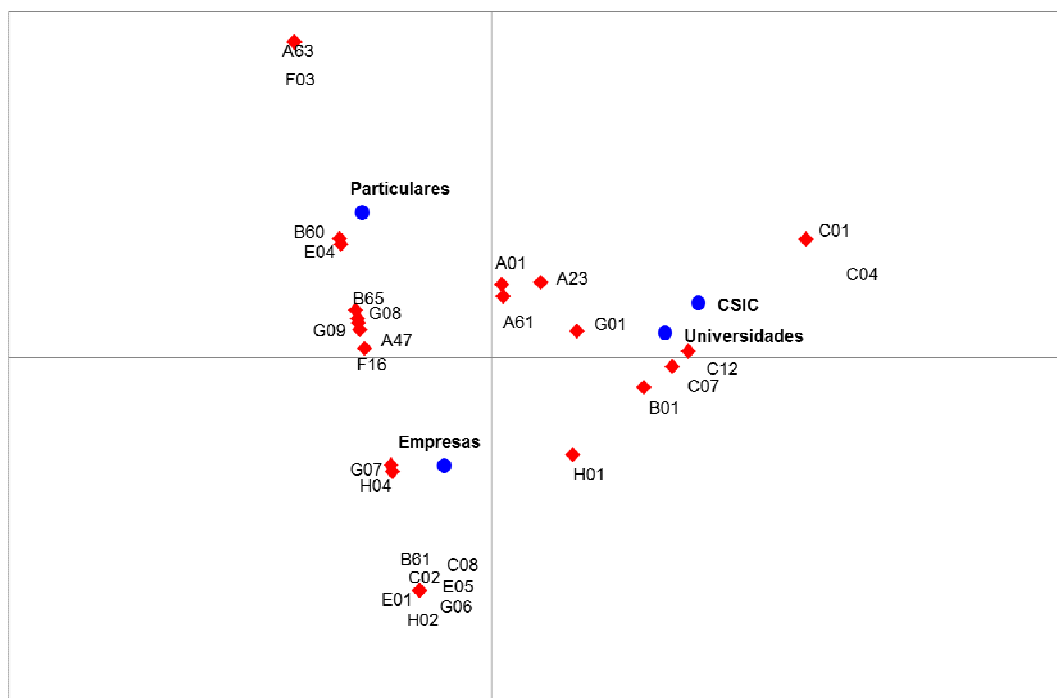


Figura 4.44: Análisis de correspondencias. Temas (nivel 3) vs sectores (Sólo se han representado los temas que aparecen al menos en 30 registros)

Nuevamente las universidades y el CSIC aparecen próximas en el mismo cuadrante y con un perfil diferenciado con respecto al resto de sectores institucionales, situándose próximos a temas en los que patentan, y que varios de ellos están relacionados con la Química, como son: C12 (Química orgánica), C7 (Bioquímica; Cerveza; Bebidas alcohólicas; Vino; Vinagre; Microbiología; Enzimología; Técnicas de mutación o de Genética), C01 (Química inorgánica), C04 (Cementos, hormigón, piedra artificial, cerámicas y refractarios) y B01 (Procedimientos o aparatos físicos o químicos en general). Algunos de estos están vinculados también con la Física (G01, Metrología).

Algo más próximos a la zona central del mapa se sitúan los particulares y empresas que muestran de nuevo un perfil amplio y difuso, ya que en su entorno se sitúan temas procedentes de distintas secciones. En el cuadrante superior izquierdo aparecen los temas más ligados a la actividad patentadora de los particulares, como B60 (Vehículos en general), E04 (Edificios), o B65 (Transporte; embalaje; almacenado; manipulación de materiales delgados o filiformes).

En la zona central del mapa se agruparían las clases en las que más patentan todas las instituciones analizadas. En este estudio, las tres clases que se sitúan en esta zona son las incluidas en la sección A (Necesidades de la Vida), todas ellas con un gran número de patentes asociadas. Se trata de Ciencias Médicas o Veterinarias e Higiene (A61), Agricultura (A01) y Alimentos (A23).

En la zona inferior izquierda se sitúa otro grupo de temas, en este caso vinculados a las patentes procedentes del ámbito empresarial, en el que destacan los dispositivos de control (G07) y Técnica de las comunicaciones eléctricas (H04).

Finalmente, como puede observarse en la Figura 4.44, aparecen 2 perfiles bien diferenciados, por un lado las universidades y el CSIC patentando en temas con un mayor componente investigador, y por otro lado las empresas y particulares con un perfil que implica, en la mayoría de los casos, investigación aplicada.

Las frecuencias absolutas de estas clases se presentan en la Tabla 4.9. A este nivel los temas de mayor producción en nuestro trabajo son las Ciencias Médicas o Veterinarias e Higiene, presentes en 435 patentes, seguidos de Metrología en 429, y Bioquímica, Bebidas Alcohólicas, Vinagre, Microbiología, Enzimología, Técnicas de Mutación o Ingeniería Genética en 340 patentes. Estos resultados coinciden con los obtenidos en el Proyecto PIPCYT en relación con las patentes de la CM solicitadas entre los años 2000 y 2007 (Gómez y otros, 2009).

Tabla 4.9: Temática de las patentes concedidas a la CM a nivel clase según código CIP

	Cód. CIP	Clase	Frec.
A	A01	Agricultura; silvicultura; cria; caza; captura; pesca	163
	A23	Alimentos o productos alimenticios; su tratamiento, no cubierto por otras clases	150
	A47	Mobiliario	142
	A61	Ciencias médicas o veterinarias; higiene	435
	A63	Deportes; juegos; distracciones	69
B	B01	Procedimientos o aparatos físicos o químicos en general	150
	B60	Vehículos en general	166
	B61	Ferrocarriles	61
	B65	Transporte; embalaje; almacenado; manipulación de materiales	131

		delgados o filiformes	
	C1	Química inorgánica	61
	C2	Tratamiento del agua, agua residual, de alcantarilla o fangos	78
	C4	Cementos: hormigón; piedra artificial; cerámicas; refractarios	69
	C07	Química orgánica	292
C	C08	Compuestos macromoleculares orgánicos; su preparación o producción química; composiciones basadas en compuestos macromoleculares	72
	C12	Bioquímica; cerveza; bebidas alcohólicas; vino; vinagre; microbiología; enzimología; técnicas de mutación o de genética	340
	E01	Construcción de carreteras, vías férreas o puentes	72
	E04	Edificios	186
E	E05	Cerraduras; llaves; accesorios de puertas o ventanas; cajas fuertes	60
	F03	Maquinas o motores de líquidos	42
F	F16	Elementos o conjuntos de tecnología; medidas generales para asegurar el buen funcionamiento de las maquinas o instalaciones; aislamiento térmico en general	71
	G01	Metrología	429
	G06	Cómputo; cálculo; conteo	130
G	G07	Dispositivos de control	155
	G08	Señalización	106
	G09	Enseñanza; criptografía; presentación; publicidad; precintos	85
	H01	Elementos o conjuntos de tecnología; medidas generales para asegurar el buen funcionamiento de las maquinas o instalaciones; aislamiento térmico en general	176
	H02	Producción, conversión o distribución de la energía eléctrica	96
	H04	Técnica de las comunicaciones eléctricas	32

4.3.2. Origen del conocimiento tecnológico de las patentes españolas

En esta sección se muestra la ponencia “*Origen del conocimiento tecnológico de las patentes españolas*” presentada en el VII Foro Evaluación de la calidad de la educación superior y la investigación, publicado en le libro de resumen de este foro. El objetivo de este trabajo era determinar el origen de la investigación que ha dado lugar a las patentes españolas.

El análisis de la innovación patentada constituye una herramienta fundamental para determinar la capacidad tecnológica de un país, región o institución. En este sentido, la procedencia del conocimiento que ha sido utilizado para generar nuevas patentes permite conocer el estado de la tecnología en los diferentes sectores de actividad

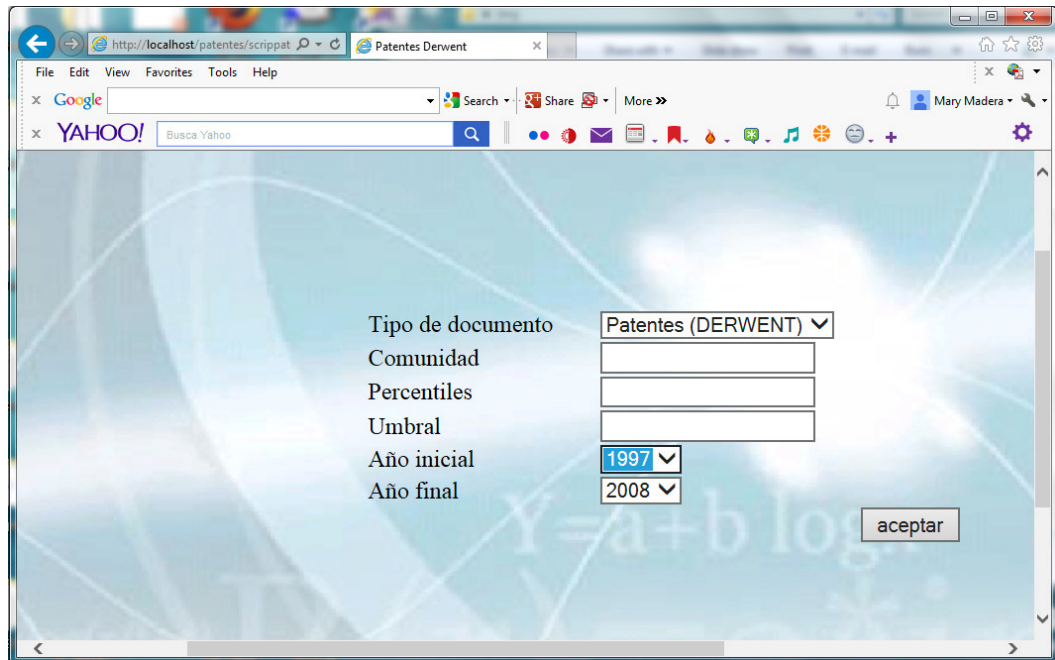
temática. Por ello, el estudio de las referencias realizadas a otras patentes nos ha permitido rastrear la tecnología a partir de la que se originan nuevas invenciones.

Los indicadores obtenidos en relación con las patentes fuentes han sido: su evolución temporal, temática (según clasificación CIP), procedencia geográfica (por Comunidades Autónomas) y el país de protección de aplicación prioritaria. Con respecto a las referencias se ha analizado: evolución temporal, nacionalidad del solicitante y temática. Los resultados figuran tanto en valores absolutos como desagregados por categorías temáticas a nivel sección de la CIP.

Mediante sendos análisis de correspondencias se establecieron las relaciones de dependencia entre los temas de las patentes fuente y los de las patentes referenciadas, así como entre los de las patentes fuente y la nacionalidad de las patentes referenciadas. El análisis de las mismas permitió establecer el índice de aislamiento (porcentaje de patentes españolas referenciadas en las patentes fuente), así como la nacionalidad del resto de las referenciadas.

Para cumplir el propósito de este estudio de caso también se utilizó el sistema desarrollado del modelo teórico-metodológico propuesto en esta investigación. A continuación se describe el proceso seguido para la obtención de los reportes que fueron utilizados para la generación de las gráficas.

Para la evaluación de las patentes fuente se utilizó la opción Patentes (*Derwent*) del módulo de reportes, en el periodo 1997 al 2008, en la Figura 4.45 se muestra esta interfaz de selección de parámetros para obtener la evolución de las patentes fuente. Como se muestra en la Figura 4.46, una vez que se seleccionan los parámetros de consulta se muestra una pantalla con una lista de reportes que pueden ser obtenidos con los parámetros seleccionados en la pantalla de la Figura 4.45; en este caso particular se deseaba obtener la evolución general de las patentes. En la Figura 4.47 se observa que se obtuvieron un total de 19.066 patentes españolas concedidas entre los años 1997 y 2008.



The screenshot shows a web browser window with the URL `http://localhost/patentes/scrippat`. The page has a light blue background with a faint mathematical formula $Y=a+b \log$. The form contains the following fields and options:

- Tipo de documento: Patentes (DERWENT) (dropdown menu)
- Comunidad: (empty text input)
- Percentiles: (empty text input)
- Umbral: (empty text input)
- Año inicial: 1997 (dropdown menu)
- Año final: 2008 (dropdown menu)
- acceptar (button)

Figura 4.45: Formulario de selección de parámetros para obtener la evolución de las patentes fuente.

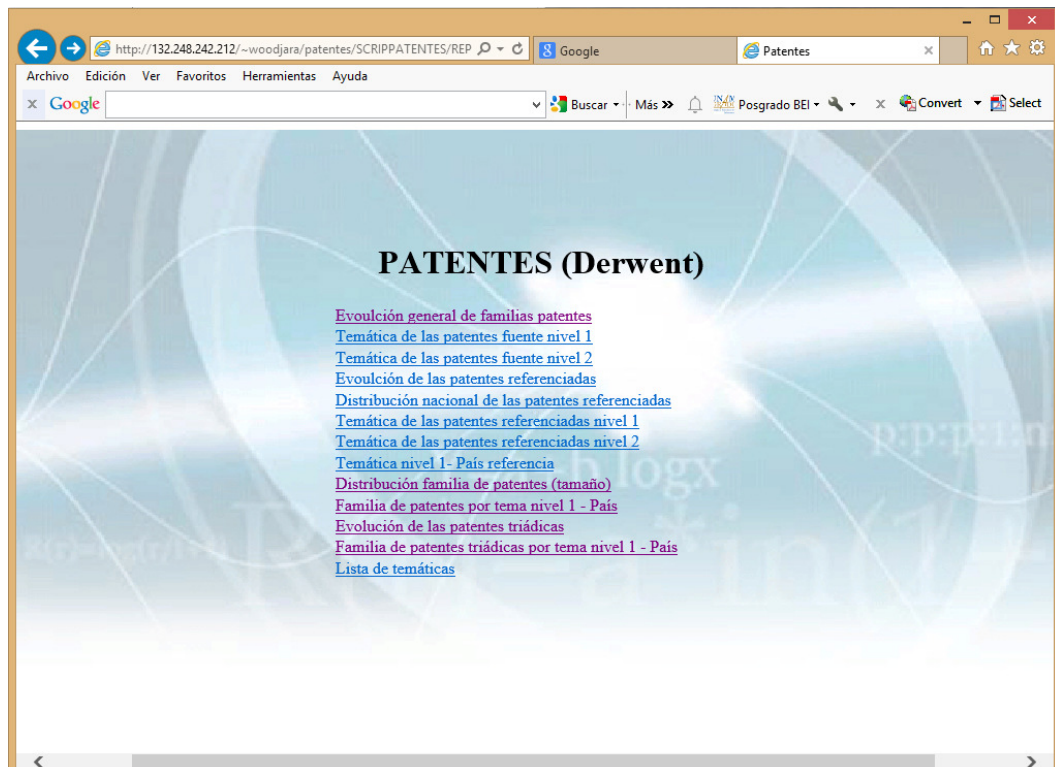


Figura 4.46: Lista de reportes de patentes *Derwent*.



Figura 4.47: Reporte de la evolución temporal de las patentes fuente.

Para el procesamiento de los datos obtenidos en los reporte se utilizó Excel y Xlstat para generar las gráficas; el usuario de este sistema lo que hace es copiar los datos del reporte generado y copiarlos en una hoja de cálculo de Excel e indicar que el delimitador de los datos es el símbolo “|” para que estos queden formateados en filas y columnas. Siguiendo este procedimiento y con las herramientas de graficación arriba mencionadas se obtiene la gráfica de la Figura 4.48 que muestra la evolución de las patentes fuente.

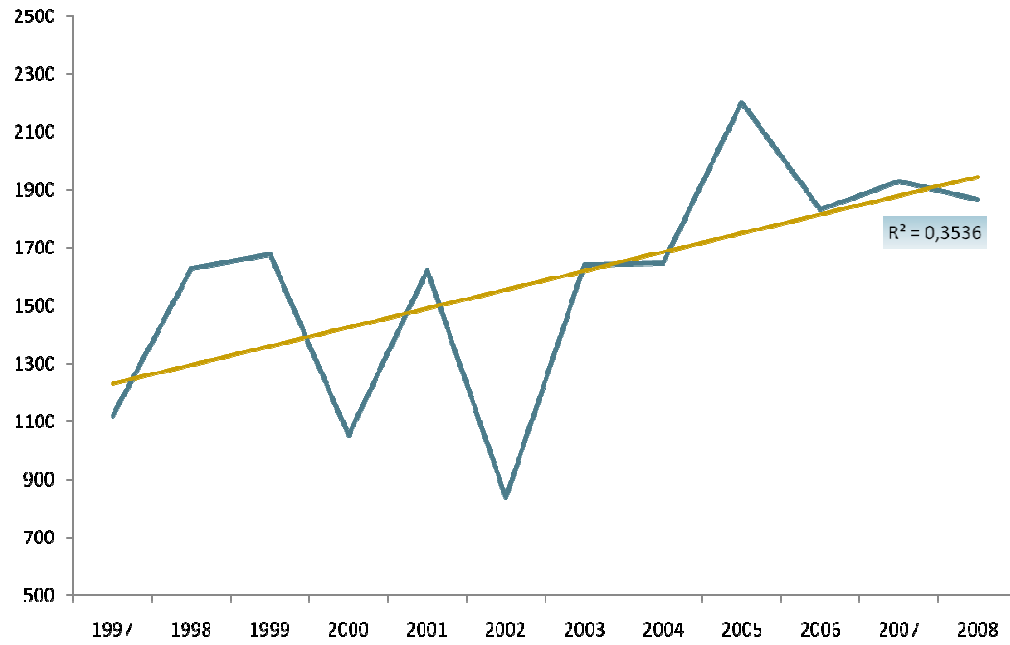


Figura 4.48: Evolución de las patentes fuente (1997-2008)

Para obtener la distribución anual de las patentes fuente y de las referenciadas (Números índice) se siguió el proceso descrito para la obtención de las patentes fuente solo que en este caso de la lista de reportes mostrados en la Figura 4.46 se seleccionó la opción “Evolución de las patentes referenciada” del cual se obtienen los datos que se observan en la Figura 4.49.

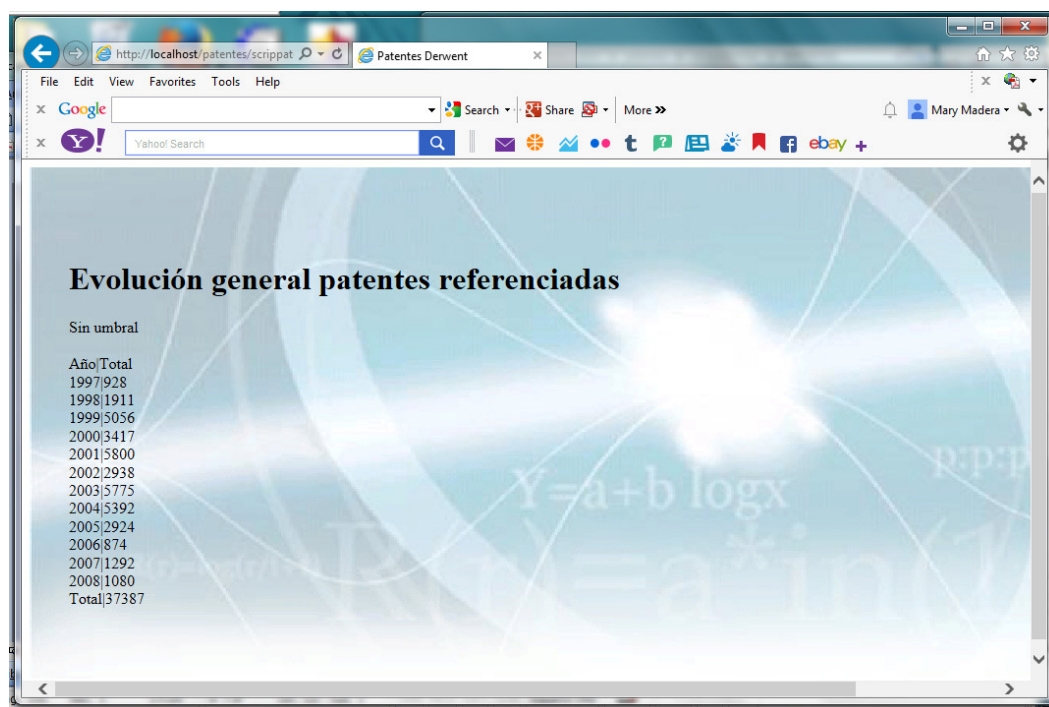


Figura 4.49: Evolución general de las patentes referenciadas.

Los datos obtenidos con el reporte de Evolución general de las patentes referenciadas son también copiados en una hoja de Excel, éstos aunados a los datos que se obtuvieron en la evolución de patentes fuente son utilizados para obtener la distribución anual de las patentes fuente y de las referenciadas. Números índice. La Tabla 4.10 muestra los números índice obtenidos con la ayuda de los reportes y en la Figura 4.50 se ve la gráfica de esta distribución.

Tabla 4.10: Distribución de patentes fuente vs. Referenciadas, núm. índice

Año	Referencias (patentes)	Número índice	Fuente (patentes)	
1997	928	100	1121	100
1998	1911	205.93	1630	145.406
1999	5056	544.83	1678	149.688
2000	3417	368.21	1053	93.934
2001	5800	625.00	1623	144.781
2002	2938	316.59	839	74.844
2003	5775	622.31	1642	146.476
2004	5392	581.03	1646	146.833
2005	2924	315.09	2202	196.432
2006	874	94.18	1834	163.604
2007	1292	139.22	1929	172.079

2008	1080	116.38	1869	166.726
Total	37387		19066	

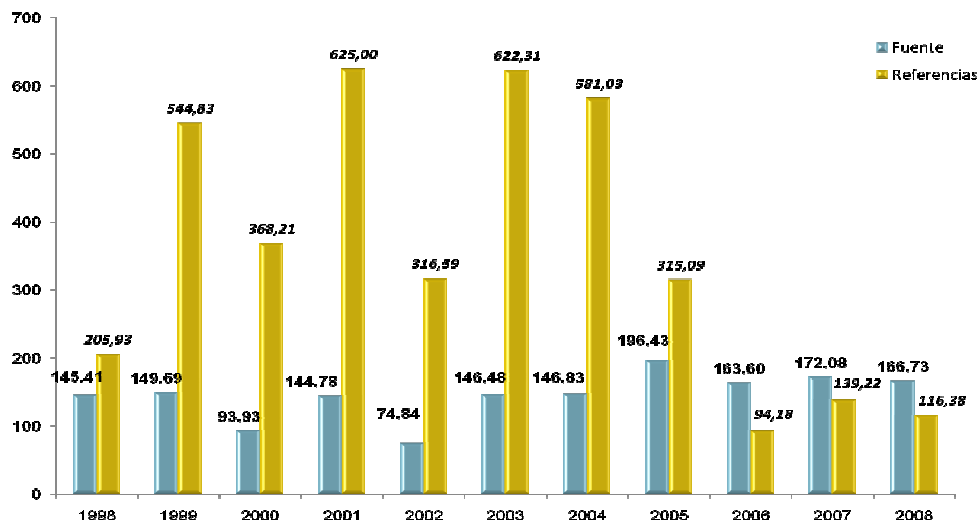


Figura 4.50: Distribución anual de las patentes fuente y de las referenciadas. Números índice.

Para generar un reporte del país de origen de las referencias a patentes se selecciona la opción “Distribución nacional de las patentes referenciadas”, este reporte genera resultados en el mismo formato de los reportes anteriores por tal motivo el proceso para llevar los resultados a Excel es el descrito anteriormente. En la Tabla 4.11 se puede apreciar un extracto del país de origen de las referencias hechas en las patentes fuente a otras patentes.

Tabla 4.11: Origen geográfico de las referencias incluidas en las patentes

País	Frecuencia	%
ESTADOS UNIDOS (US)	6458	29,25
ESPAÑA (ES)	3781	17,12
ALEMANIA (DE)	2966	13,43
FRANCIA (FR)	2670	12,09
JAPÓN (JP)	1195	5,41
REINO UNIDO (GB)	1159	5,25
ITALIA (IT)	519	2,35

SUIZA (CH)	459	2,08
CANADÁ (CA)	388	1,76
SUECIA (SE)	311	1,41
AUSTRALIA (AU)	243	1,10
PAISES BAJOS (NL)	225	1,02
UNIÓN SOVIÉTICA (SU)	193	0,87
FINLANDIA (FI)	165	0,75
AUSTRIA (AT)	139	0,63
TAIWAN (TW)	130	0,59
ISRAEL (IL)	120	0,54
DINAMARCA (DK)	116	0,53
TOTAL	22081	

Para realizar el gráfico de dispersión entre la temática de las patentes fuentes y de las patentes referenciadas a nivel sección de la CIP se deben utilizar las opciones “Temática de las patentes fuente nivel 1” y “Temática de las patentes referenciadas nivel 1” de la lista de reportes de la Figura 4.46.

Debido a que el formato de los reportes es igual de aquí en adelante solo se mostraran las tablas con los datos de éstos en Excel, así como las operaciones que sobre ellos realiza el usuario encargado de obtener las gráficas.

En la Tabla 4.12 se ven el código de las temáticas a nivel sección de la CIP de las patentes fuente, la descripción y el total de patentes en cada una de estas temáticas; la columna del porcentaje de cada temática con relación al total fue calculado por el usuario. En la Tabla 4.13 se muestran este mismo tipo de resultados pero son datos relacionados con las temáticas de las patentes referenciadas.

Tabla 4.12: Distribución de la temática de las patentes fuente a nivel sección

Código Temática	Descripción	Total	%
B	Técnicas industria / Transportes	5871	23.67
A	Necesidades de la vida	5643	22.75
G	Física	3139	12.65

C	Química / metal	2978	12.01
E	Construcción	2355	9.49
F	Mecánica	2328	9.39
H	Electricidad	2031	8.19
D	Textiles / Papel	460	1.85
Total		24805	100.00

Tabla 4.13: Distribución de la temática de las patentes referenciadas a nivel sección

Código temática	Descripción	Total	%
B	Técnicas industria / Transportes	12952	24.62
A	Necesidades de la vida	11368	21.61
G	Física	6855	13.03
F	Mecánica	5750	10.93
E	Construcción	5055	9.61
C	Química / metal	4842	9.20
H	Electricidad	4801	9.13
D	Textiles / Papel	989	1.88
Total		52612	100

Para graficar la dispersión de la temática de las patentes fuentes vs. la temática de las patentes referenciadas presentada en la Figura 4.51 el usuario solo tuvo que combinar los datos de de las Tablas 4.12 y 4.13 y obtener la diferencia entre los porcentajes. Esta unión de datos se muestra en la Tabla 4.14.

Tabla 4.14: Temática patentes fuente vs referenciadas

Código temática	Descripción	% Patentes fuente	% Patentes referencia	Diferencia
B	Técnicas industria / Transportes	23.67	24.62	-0.95
A	Necesidades de la vida	22.75	21.61	1.14
G	Física	12.65	13.03	-0.37
F	Mecánica	12.01	10.93	1.08
E	Construcción	9.49	9.61	-0.11
C	Química / metal	9.39	9.20	0.18
H	Electricidad	8.19	9.13	-0.94
D	Textiles / Papel	1.85	1.88	-0.03

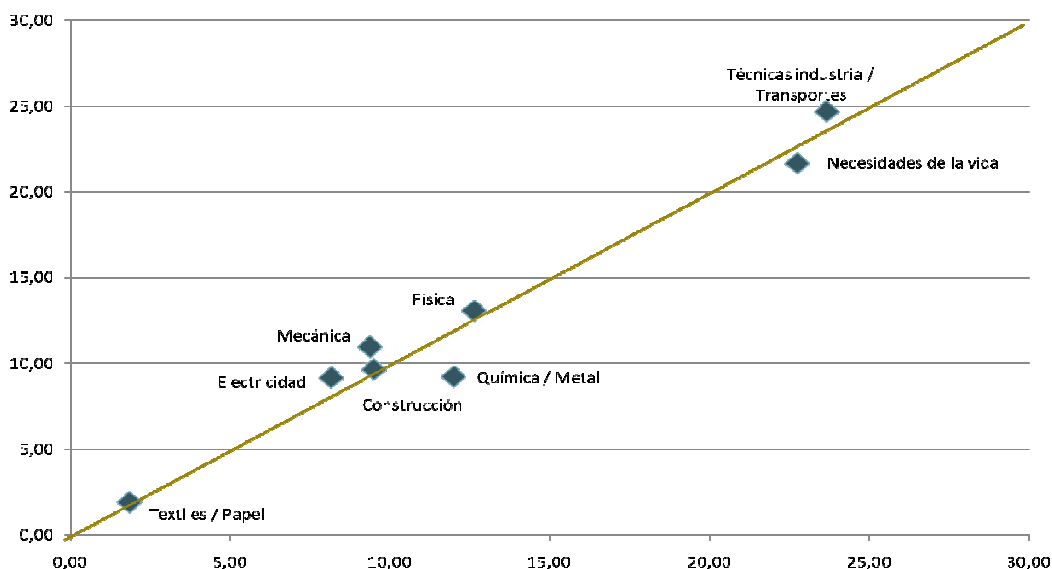


Figura 4.51: Temática de las patentes fuentes y de las patentes referenciadas. Nivel sección de la CIP.

El proceso para obtener los datos que serán utilizados en la graficación de la dispersión entre la temática de las patentes fuentes y de las patentes referenciadas a nivel clase de la CIP es igual al descrito en el caso de la dispersión a nivel sección solo que las opciones de la lista de reportes de la Figura 4.46 que se deben utilizar son la “Temática de las patentes fuente nivel 2” y la “Temática de las patentes referenciadas nivel 2”. No obstante que el reporte incluye todas las temáticas presentes en cada una de las patentes y dado que la lista de temáticas reportadas es muy larga, al momento de utilizar los datos para graficarlos entra en juego la experiencia del usuario ya que selecciona las temáticas más representativas puesto que incluirla todas solo ocasionaría “ruido” en la gráfica. En la Tabla 4.15 se muestra un segmento de las temáticas mas representativas seleccionadas por el usuario y en la Figura 4.52 se ve la dispersión entre las temáticas de las patentes fuente y las referenciadas más representativas a nivel clase.

Tabla 4.15: Temática patentes fuente vs referenciadas a nivel clase.

Código	Descripción	Total fuente	% Fuente	Total ref.	% ref.	% Diferencia
A01	AGRICULTURA; SILVICULTURA; CRIA; CAZA; CAPTURA; PESCA	986	3,42	2039	3,27	0,15
A23	ALIMENTOS O PRODUCTOS ALIMENTICIOS; SU TRATAMIENTO, NO CUBIERTO POR OTRAS CLASES	861	2,98	1436	2,30	0,68

A47	MOBILIARIO	935	3,24	2192	3,52	-0,27
A61	CIENCIAS MÉDICAS O VETERINARIAS; HIGIENE	2103	7,29	3868	6,20	1,09
A63	DEPORTES; JUEGOS; DISTRACCIONES	366	1,27	883	1,42	-0,15
B01	PROCEDIMIENTOS O APARATOS FÍSICOS O QUÍMICOS EN GENERAL	598	2,07	1456	2,34	-0,26
B29	TRABAJO DE LAS MATERIAS PLÁSTICAS; TRABAJO DE SUSTANCIAS EN ESTADO PLÁSTICO EN GENERAL	385	1,33	1000	1,60	-0,27
B60	VEHÍCULOS EN GENERAL	921	3,19	2205	3,54	-0,34
B62	VEHÍCULOS TERRESTRES QUE SE DESPLAZAN DE OTRO MODO QUE POR RAILES	309	1,07	694	1,11	-0,04
B65	TRANSPORTE; EMBALAJE; ALMACENADO; MANIPULACION DE MATERIALES DELGADOS O FILIFORMES	1445	5,01	3031	4,86	0,15
C07	QUÍMICA ORGÁNICA	918	3,18	1080	1,73	1,45
C12	BIOQUÍMICA; CERVEZA; BEBIDAS ALCOHÓLICAS; VINO; VINAGRE; MICROBIOLOGÍA; ENZIMOLOGÍA; TÉCNICAS DE MUTACIÓN O DE GENÉTICA	735	2,55	823	1,32	1,23

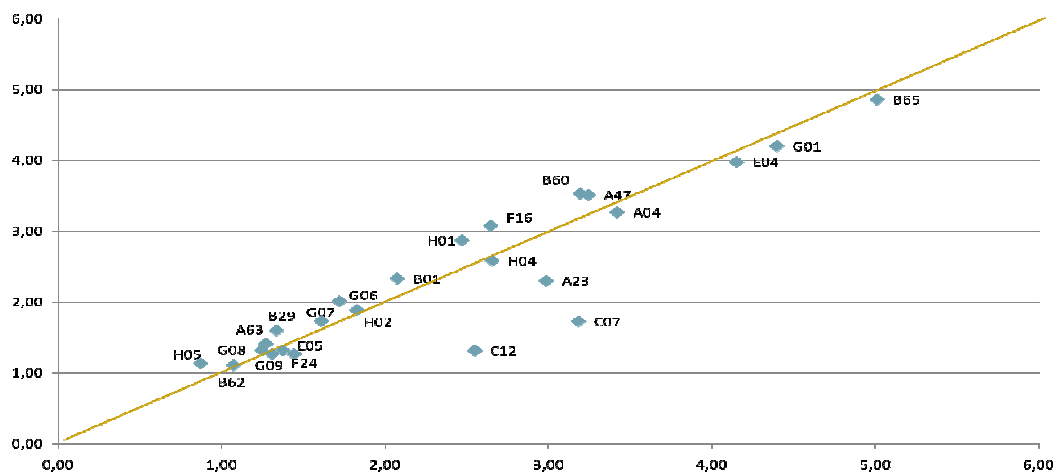


Figura 4.52: Temas de las patentes fuentes y de las patentes referenciadas. Nivel clase de la CIP.

Para concluir con los resultados presentados en esta sección en la Figura 4.53 se muestra el análisis de correspondencia entre la temática a primer nivel de las patentes fuente y el país de procedencia de las patentes referenciadas.

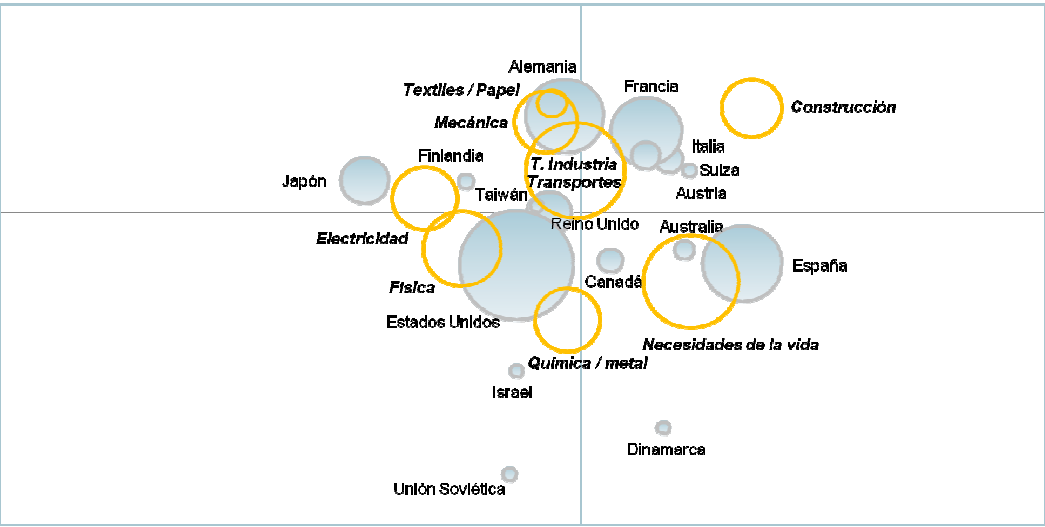


Figura 4.53: Temas vs. Países de las patentes referenciadas. AC

Esta grafica se obtiene seleccionando la opción “Temática nivel 1 – país de referencia”, ésta genera una matriz de contingencia en la cual se tienen los códigos de los países de las patentes referenciadas en el renglón y los códigos de las temáticas a nivel 1 en la columna, el cruce de renglón-columna representa el total de patentes fuente con temática “X” que han hecho referencia a patentes del país “Y”. En la Tabla 4.16 se muestra la matriz generada por el reporte seleccionado, ésta es usada por Xlstat para generar el análisis de correspondencia de la temática vs los países de las patentes referenciadas (Figura 4.49).

Tabla 4.16: Temática patentes fuente vs. país de patente referenciada

Temática/País	AT	AU	CA	CH	DE	DK	ES	FI	FR	GB	IL	IT	JP	SU	TW	US	Total
A	43	97	118	136	740	62	1466	45	829	333	45	178	212	52	47	2065	6468
B	45	65	118	168	1179	28	1026	48	940	343	25	197	449	22	32	2185	6870
C	25	38	79	70	390	33	533	15	232	169	26	50	179	45	6	1109	2999
D	2	1	11	16	130	4	80	8	46	35	3	42	49	2	5	131	565
E	32	43	45	78	465	3	550	22	492	145	9	58	83	9	14	497	2545
F	20	32	46	68	562	15	307	16	430	171	11	60	181	8	15	845	2787
G	21	47	69	68	568	17	460	45	490	296	40	67	355	52	31	1615	4241
H	17	28	41	39	465	11	298	45	332	159	18	57	329	17	28	1087	2971
Total	205	351	527	643	4499	173	4720	244	3791	1651	177	709	1837	207	178	9534	

4.3.3. Interés de las invenciones españolas a partir del estudio de las familias de patentes

Los indicadores obtenidos a partir del análisis de patentes constituyen una herramienta de gran interés para valorar la capacidad tecnológica de un país, región o institución. Dentro de éste ámbito, el estudio de las familias de patentes (entendidas como los distintos documentos de patentes a los que una misma invención da lugar al ser protegida en distintos países) sirve de apoyo a investigaciones tanto de tipo económico, estadístico y cuantitativo, como la que aquí se presenta. Los datos sobre tamaño de las familias se han usado para extraer inferencias sobre el valor de las invenciones ya que a los titulares de las mismas les resulta muy costoso mantener su protección durante periodos largos y en distintos países, por lo que el tamaño se puede vincular con la importancia que tienen las patentes para los sectores productivos.

El propósito de éste trabajo consiste en determinar las características de las familias de las patentes prioritarias concedidas por la OEPM a invenciones con solicitante de nacionalidad española durante el periodo 1997-2008. Uno de los objetivos concretos es establecer la relación existente entre el tamaño de las familias y los temas de las invenciones, así como entre el tamaño y los estados designados (países y oficinas) donde se protegen.

Al igual que en el estudio de caso anterior, para caracterizar el interés de las invenciones españolas a partir del estudio de las familias de patentes se utilizó la opción “Patentes (*Derwent*)” del módulo de reportes mostrado en la Figura 4.45, el periodo estudiado fue del año 1997 al 2008, de la lista de reportes mostrada en la Figura 4.45 se seleccionó el reporte “Evolución de familia de patentes”.

Para el procesamiento de los datos obtenidos en los reportes también se utilizó Excel y Xlstat para generar las gráficas. En la Tabla 4.17 se visualiza la evolución temporal de las patentes prioritarias obtenidas mediante el reporte seleccionado, ésta es la fuente de datos para generar la gráfica de la Figura 4.54.

Tabla 4.17: Evolución temporal de las patentes prioritarias.

Año	# patentes prioritarias
1997	1108
1998	1614
1999	1680
2000	1055
2001	1635
2002	856
2003	1647
2004	1659
2005	2208
2006	1833
2007	1940
2008	1892
Total	19127

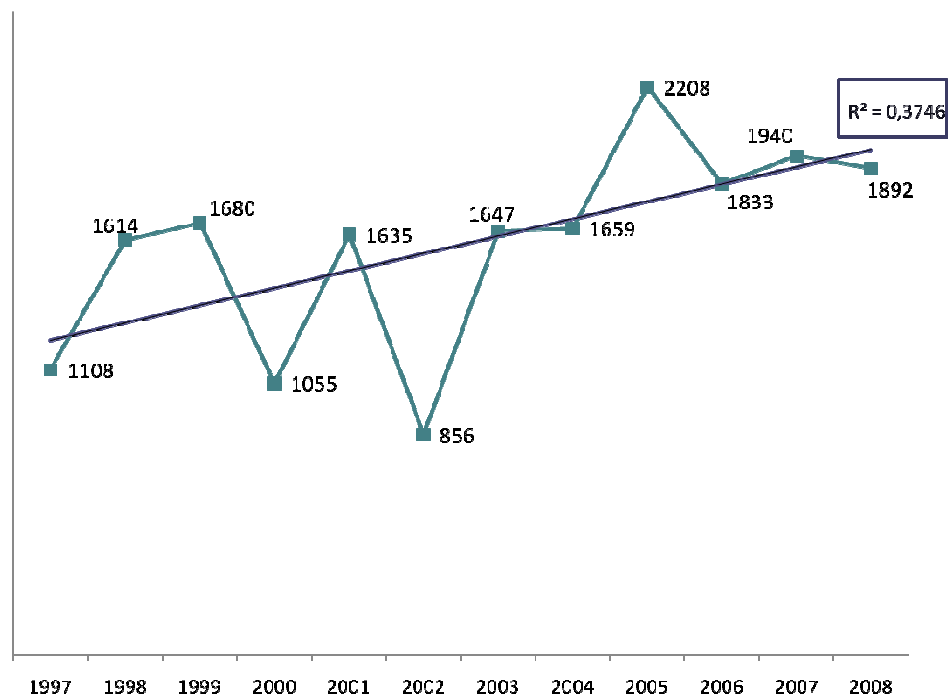


Figura 4.54: Evolución temporal patentes prioritarias

En la Tabla 4.18 se muestran la distribución de las familias de patentes, para obtener esta distribución se usó la opción “Distribución de familia de patentes” la cual genera

un reporte que agrupa las familias de patentes por el número de miembros que las conforman. En la misma tabla se indica la frecuencia, dato que es calculado por el usuario.

Tabla 4.18: Tamaño de las familias de patentes

Nº de miembros	Nº de familias	%	Nº de miembros	Nº de familias	%
1	325	1,699	22	20	0,105
2	12411	64,887	23	19	0,099
3	1419	7,419	24	12	0,063
4	984	5,145	25	14	0,073
5	914	4,779	26	10	0,052
6	586	3,064	27	12	0,063
7	527	2,755	28	3	0,016
8	398	2,081	29	11	0,058
9	289	1,511	30	8	0,042
10	226	1,182	31	4	0,021
11	205	1,072	32	4	0,021
12	155	0,810	33	4	0,021
13	107	0,559	34	3	0,016
14	125	0,654	35	1	0,005
15	89	0,465	36	4	0,021
16	57	0,298	38	3	0,016
17	55	0,288	41	1	0,005
18	41	0,214	44	1	0,005
19	38	0,199	47	2	0,010
20	25	0,131	53	1	0,005
21	13	0,068	59	1	0,005
			Total	19127	100

En la Figura 4.55 se visualiza la distribución de la temática a primer nivel (CIP) de las familias de patentes, para obtener esta visualización se usa la opciones “Temática de las patentes fuente nivel 1” de la lista de reportes de la Figura 4.46. En la Tabla 4.19 se ven el código de las temáticas a nivel sección de la CIP de las familias de patentes, la descripción y el total de familias de patentes en cada una de estas temáticas; la

columna de la frecuencia de cada temática con relación al total fue calculado por el usuario.

Tabla 4.19: Distribución de la temática de las familias de patentes a nivel sección

Temática fam. de patentes	Descripción	Total	Frecuencia
B	Técnicas industria / Transportes	5871	23.67
A	Necesidades de la vida	5643	22.75
G	Física	3139	12.65
C	Química / metal	2978	12.01
E	Construcción	2355	9.49
F	Mecánica	2328	9.39
H	Electricidad	2031	8.19
D	Textiles / Papel	460	1.85
Total		24805	100.00

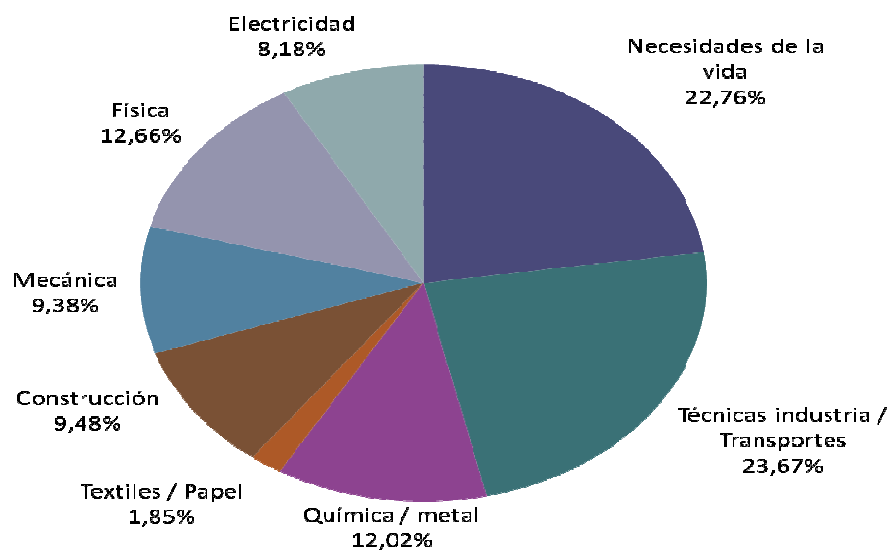


Figura 4.55: Temática a primer nivel (CIP) de las familias de patentes

En la Tabla 4.20 se muestran la distribución de las familias de patentes por tema a nivel sección (CIP) y por país de protección, esta distribución se obtiene usando la opción “Familia de patentes por tema nivel 1- País”. El reporte generado mediante esta

opción muestra más datos de los utilizados en la tabla, pero nuevamente aquí juega un papel importante la experiencia de la persona que está usando los datos ya que en este caso, del total de países, solamente utilizo aquellos en los cuales el total de la última fila era 1000 ya que considero eran los más representativos y de esta forma evitaba introducir “ruido” en el análisis de correspondencia de la Figura 4.56.

Tabla 4.20: Distribución de la temática de las familias de patentes a nivel sección vs País de protección

Temática/ País	AU	BR	CN	DE	EP	ES	JP	MX	US	WO	Total
Necesidades de la vida	849	334	327	613	1333	5757	517	388	889	1442	12449
Técnicas industria / Transportes	629	251	268	767	1340	5978	365	237	789	1142	11766
Química / metal	639	249	251	428	892	3104	472	281	639	1053	8008
Textiles / Papel	51	25	36	86	172	471	25	28	85	107	1086
Construcción	159	75	76	202	404	2402	65	82	172	337	3974
Mecánica	221	104	126	326	589	2349	122	78	326	439	4680
Física	362	129	109	264	618	3244	174	117	362	671	6050
Electricidad	218	133	134	227	548	2103	152	121	299	421	4356
Total	3128	1300	1327	2913	5896	25408	1892	1332	3561	5612	52369

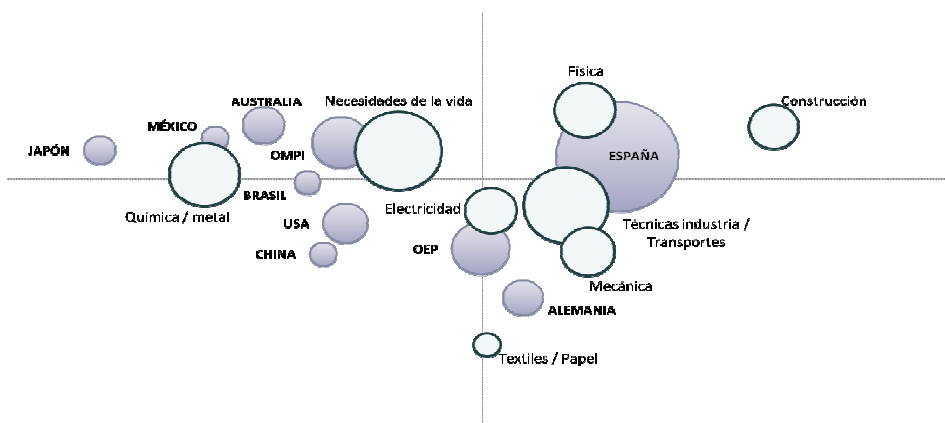


Figura 4.56: Temática de las familias de patentes vs País

Para obtener la distribución de las familias de patentes triádicas por tema a nivel sección (CIP) y por país de protección se usa la opción “Familia de patentes triádicas por tema nivel 1- País”, en este caso se debe llenar un parámetro mas en el formulario de la Figura 4.45, este parámetro es el umbral ya que solamente se van a seleccionar las familias de patentes que tengan 20 o más miembros; la tabla de resultados es igual a la que se obtiene en la opción “Familia de patentes por tema nivel 1- País”. La gráfica de la Figura 4.57 que muestra las relaciones entre las familias de patentes de las temáticas, se obtuvo mediante la herramienta de visualización *Pajek*, para generarla se transformaron los datos a un fichero “.net” en el formato requerido por esta herramienta. Es importante mencionar que los ficheros en formato “.net” también son usados como fuente de entrada en las herramientas de visualización *Gephi* y *VOSviewer*.

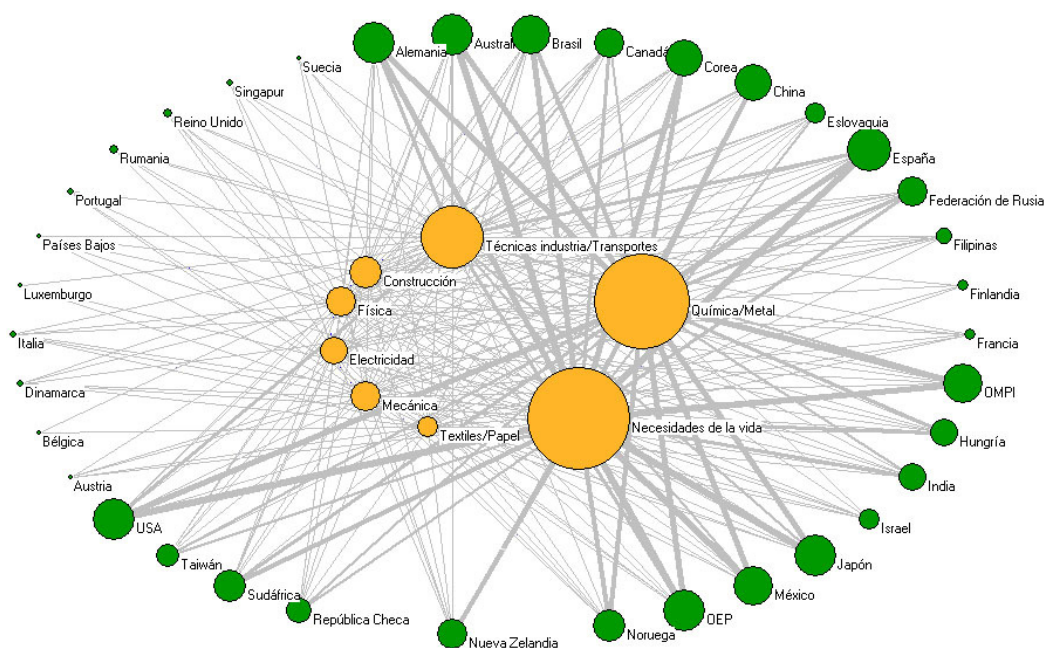


Figura 4.57: Temática de las familias de patentes vs países de las familias de patentes (20 o más miembros)

En la Tabla 4.21 se muestra la evolución temporal de las familias de patentes triádicas, datos que son utilizados como fuente para generar la gráfica de barras de la Figura 4.58. Este reporte se obtiene mediante la opción “Evolución de patentes triádicas”.

Tabla 4.21: Evolución temporal de las familias de patentes triádicas

Año	# familias de patentes triádicas
1997	1108
1998	1614
1999	1680
2000	1055
2001	1635
2002	856
2003	1647
2004	1659
2005	2208
2006	1833
2007	1940
2008	1892
Total	19127

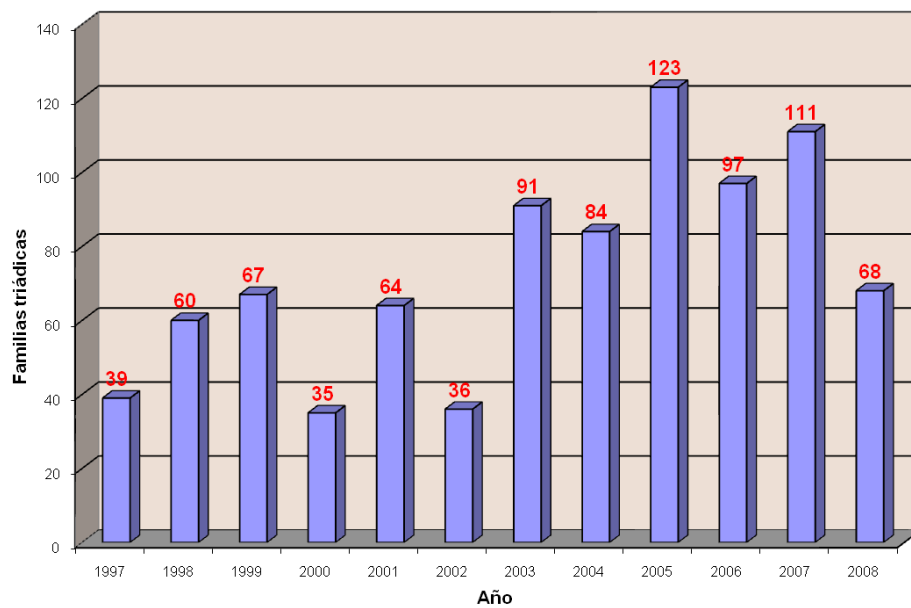


Figura 4.58: Evolución temporal de las familias de patentes triádicas

En la Tabla 4.22 se muestra la distribución de las familias de patentes triádicas por código de las temáticas a nivel sección de la CIP, la descripción y el total de familias

triádicas en cada una de estas temáticas; la columna del frecuencia de cada temática con relación al total fue calculado por el usuario.

Tabla 4.22: Temática a primer nivel (CIP) de las familias de patentes triádicas

Temática Primer nivel (CIP)	# familias triádicas	Frecuencia
Necesidades de la vida	442	28,6
Química/Metal	383	24,8
Técnicas industria/Transportes	291	18,9
Física	134	8,7
Electricidad	121	7,9
Mecánica	97	6,3
Construcción	53	3,4
Textiles/Papel	22	1,4
Total	1543	100

CAPÍTULO 5

CONCLUSIONES

5. Conclusiones

1. El modelo teórico-Metodológico diseñado en esta investigación, así como su implementación a través de una aplicación *Web* y su comprobación con el estudio de caso de patentes españolas concedidas en el periodo 1996-2008, demuestran la factibilidad en el uso de estos resultados en la generación de indicadores bibliométricos sobre patentes.
2. La bases de datos *Interpat* proporciona un conjunto de datos de gran importancia para el estudio de las patentes, aunque tiene el inconveniente que los devuelve en un archivo con formato del procesador de palabras *Word*, eso aunado a que a lo largo del desarrollo de este proyecto el etiquetado de los datos cambiaba, ha dificultado su procesamiento para su utilización en la obtención de indicadores bibliométricos sobre patentes.
3. La base de datos *Derwent Innovations Index*, posee una interfaz de consulta que proporciona varios formatos para descargar los datos bibliográficos de las patentes lo cual permite que el programador elija el formato que más le facilite el procesamiento de éstos.
4. Los datos de la base *Worldwide-espacenet*, se descargan en formato *CSV*, formato que facilita en gran medida el procesamiento de datos, aunque resulta conveniente señalar que la interfaz de consulta de esta base de datos tiene como desventaja que solo permite hacer las búsqueda de 10 números de patentes por consulta.
5. Las tres bases de datos analizadas en este trabajo han mostrado tener una gran complementariedad en la información que incluyen, lo cual ha sido de gran interés, ya que ha proporcionado una batería de indicadores relevantes y muy completa para el análisis de patentes.
6. El modelo teórico-metodológico de integración de las bases de datos bibliográficas de patentes propuesto toma como fuente de datos principal los

datos obtenidos de la base de datos nacional debido a que este tipo de bases proporcionan datos más específicos y actualizados que las bases de datos internacionales.

7. La base de datos *Derwent Innovation Index* tiene los números de patente que son referenciadas por los inventores y los examinadores, estas referencias a patentes se consideran de gran importancia para obtener información sobre los nexos informativos de las patentes. Debido a la granularidad de los datos, aunado al formato CSV que brinda la base de datos *Worldwide-espacente* fue una de las causas primordiales para decidir el uso de esta base para extraer los datos de las patentes referenciadas por las patentes fuente.
8. La extracción y procesamiento de los registros de las patentes de las tres bases de datos se ha realizado en cascada, es decir, las patentes de la base de datos nacional sirven como datos de entrada en la búsqueda de las patentes en *Derwent Innovation Index*, y son los datos de esta última los de entrada en la base *Worldwide-espacenet*. Es importante hacer énfasis en que el número de patente es el atributo que sirve de unión entre las tres bases de datos.
9. En la base de datos diseñada se consideraron todas las etiquetas de las tres bases de datos para que permita almacenar e integrar todos los datos extraídos de estas bases.
10. Con fundamento en los indicadores bibliométricos y los indicadores que se describen en el *Manual de estadísticas de patentes de la OCDE* se desarrollaron los *scripts* que permiten el acceso al repositorio que albergan los datos de patentes para generar reportes y de esta forma obtener indicadores que pueden ser de gran apoyo a los responsables de la toma de decisiones.
11. Con el estudio de caso “*Capacidad innovadora de la Comunidad de Madrid a partir de las patentes concedidas entre 1996 y 2007*” se demuestra que el modelo propuesto permite hacer la caracterización de una región específica, esto se debe a que es posible utilizar el subconjunto de datos de la oficina de patentes nacional

debido a que esta base proporciona datos más específicos de la región que se analiza.

12. Al igual que en estudio de caso anterior, en el correspondiente al “*Interés de las invenciones españolas a partir del estudio de las familias de patentes*” se utilizan el subconjunto de datos proporcionados por *Derwent Innovation Index*. El gran interés de ésta radica en que es la única de las tres bases que tiene familias de patentes.
13. En el estudio de caso “*Origen del conocimiento tecnológico de las patentes españolas*” se usan los datos resultantes de la integración de las tres bases de datos, se puede observar que al integrar estos datos se pueden hacer cruces entre más variables y de esta forma obtener mucha más información.
14. Con el modelo propuesto es posible utilizar subconjuntos de datos, es decir, usar solamente los datos de una de las bases de datos fuente y obtener reportes con indicadores de datos específicos, pero al integrarlas y usarlas en conjunto se proporciona más riqueza a la información.
15. Por último, se considera importante mencionar que, aunque durante la fase de análisis de la aplicación *Web* desarrollada se trabajó con la base de datos de la Oficina Española de Patentes y Marcas debido a que el modelo metodológico se probó con un conjunto de datos bibliográficos de patentes españolas, este módulo puede ser desarrollado en función de los datos fuente con que cuente la oficina de propiedad intelectual del país que se va a abalizar.

REFERENCIAS

Referencias

- ACOSTA M, CORONADO D. (2003). Science-technology flow in spanish regions. An analysis of scientific citations in patents. *Research Policy*. (32): 1783–1803.
- ACOSTA M, CORONADO D, FERNÁNDEZ A. (2009). Exploring the quality of environmental technology in Europe: evidence from patent citations. *Scientometrics*. 80(1): 131–52.
- AL-HAWAMDEH S. (2002). Knowledge management: re-thinking information management and facing the challenge of managing tacit knowledge. *Information Research*. 8(1): 143.
- ALBERT A, GRANADINO BN, PLAZA LM. (2007) Scientific and technological performance evaluation of the spanish council for scientific research (CSIC) in the field of biotechnology. *Scientometrics*. 70(1): 41–51.
- ALBERT A, PLAZA LM. (2004). The transfer of knowledge from the spanish public R&D system to the productive sectors in the field of biotechnology. *Scientometrics*. 59(1): 3–14.
- ÁLVAREZ GARCÍA JC. (2010). *Clasificación automática de información en portales Web mediante técnicas de clustering* [tesis doctoral]. Salamanca: Departamento de Informática y Automática. Universidad de Salamanca.
- ARVANITIS A, RUSSELL J, ROSAS A. (1996). Experiences with the national citation reports database for measuring national performance: the case of Mexico. *Scientometrics*. 35(2): 247–55.
- ATALLAH G, RODRÍGUEZ G. (2006). Indirect patent citations. *Scientometrics*. 67(3): 437–65.
- AUSTIN, R. (2001). the complete markush structure search: Mission impossible? En *PIUG North East Workshop*. Germany: Fachinformationszentrum Karlsruhe. 1–78.

- AYUSO SÁNCHEZ M, AYUSO GARCÍA M. (2003). Revisión de los estudios orientados a la medición de las capacidades tecnológicas por medio de la literatura patente. Propuesta de análisis estadístico y evaluación de la calidad de una base de datos de patentes. *Revista General de Información y Documentación*. 13(1): 151–172.
- AZAGRA-CARO JM, ARCHONTAKIS F, YEGROS-YEGROS A. (2007). In which regions do universities patent and publish more?. *Scientometrics*. 70(2): 251–66.
- AZAGRA-CARO JM *et al.* (2009). What do patent examiner inserted citations indicate for a region with low absorptive capacity?. *Scientometrics*. 80(2): 441–55.
- AZAGRA-CARO JM, ROMERO-de PABLOS A. (2009). Los determinantes institucionales de las patentes del consejo superior de investigaciones científicas: una aproximación histórica y una dialéctica con la economía. *Revista Española de Documentación Científica*. 32(2): 9–33.
- AZAGRA CARO JM, YEGROS YEGROS A, ARCHONTAKIS F. (2006). What do university patent routes indicate at regional level?. *Scientometrics*. 66(1): 219–30.
- BALESTRI M, MANGIARACINA S, NOBILI D. (2001). Bibliometric S&T indicators to comply with users' needs. *Research Evaluation*. 10(1): 5–12.
- BARADOL KA, KUMBAR SS. (1998). Interdisciplinary nature of library science. *Annals of Library Science and Documentation*. 45(2): 49–56.
- BARROSO W, QUONIAM L, PACHECO E. (2009). Patents as technological information in Latin America. *World Patent Information*. (31): 207–15.
- BASTIAN M, HEYMANN S, JACOMY M. (2009). Gephi: an open source software for exploring and manipulating networks. En *International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence. Disponible en: <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154> [Consultado: 10 de marzo de 2012].

- BELKHODJA O, LANDRY R. (2007). The triple-helix collaboration: why do researchers collaborate with industry and the government? what are the factors that influence the perceived barriers?. *Scientometrics*. 70(2): 301–32.
- BHATTACHARYA S, KRETSHMER H, MEYER M. (2003). Characterizing intellectual space between science and technology. *Scientometrics*. 58(2): 369–390.
- BLANK-EDELMAN DN. *Perl for system administration*. (2000). United States of America: O'reilly and Associates.
- BLOCH C. (2006). Innovation measurement: present and future challenges. En *EUROSTAT Conference "Knowledge economy: challenges for Measurement"*. Luxemburgo: European Communities.
- BOOTH UDALL FULLER, Divisional Patent Applications: Patent Application Types and Strategies – Part 7, disponible en: <http://boothudall.com/articles/other-nonprovisional-patent-applications-patent-application-types-and-strategies-part-5> [Consultado: 10 de marzo de 2012].
- BONILLA CALERO, AI. (2009). *La colaboración y la visibilidad en las disciplinas de física en Science Citation Index y arXiv (2000-2005)* [tesis doctoral]. Getafe: Departamento de Biblioteconomía y Documentación. Universidad Carlos III de Madrid.
- BORDONS M. (2001). Aspectos metodológicos en la obtención de indicadores bibliométricos. *Cuadernos de Indicios*. 1: 17–26.
- BOTOLF MAURSETH P, VERSPAGEN B. (2002). Knowledge spillovers in Europe: a patent citation analysis. *Scandinavian Journal of Economics*. 104(4): 531–545.
- CALLAERT J *et al.* (2006). Traces of prior art: an analysis of non-patent references found in patent documents. *Scientometrics*. 69(1); 3–20.
- CAVALLER V. (2009). Scientometrics and patent bibliometrics in R&D analysis. A new approach to valuation of intangible assets. *The Journal of Information and Knowledge Management Systems*. 39(1): 80–91.

- COMAI A, TENA J, VERGARA JC. (2006). Software para la vigilancia tecnológica de patentes: evaluación desde la perspectiva de los usuarios. *El Profesional de la Información*. 15(6): 452–58.
- COSTA RIBEIRO L *et al.* (2010). Matrices of science and technology interactions and patterns of structured growth: implications for development. *Scientometrics*. 83(1): 55–75.
- COZENS, S, WAINWRIGHT P. (2000). *Beginning Perl*. United States of America: Wrox Press.
- CRISCUOLO P. (2006). The ‘home advantage’ effect and patent families. a comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*. 66(1): 23–41.
- CRISCUOLO P, VERSPAGEN B. (2008). Does it matter where patent citations come from? inventor vs. examiner citations in european patents. *Research Policy*. (37): 1892–1908.
- CUELLEN SE. (2009). A practical approach to the reformed IPC. *World Patent Information*. (31): 193–98.
- CZARNITZKI D, GLÄNZEL W, HUSSINGER K. (2009). Heterogeneity of patenting activity and its implications for scientific research. *Research Policy*. (38): 26–34.
- DENG Y. (2007). Private value of european patents. *European Economic Review*. (51): 1785–1812.
- DENT C *et al.* (2006). Research use of patented knowledge: a review. En *OECD Science, Technology and Industry Working Papers, 2006/2*. United States of America: OECD Publishing. 1-53.
- DERNIS H, GUELLEC D, VAN POTTELSBERGHE B. (2001). Using patents counts for cross-country comparisons of thechnology output. En *OECD Science, Technology and Industry, STI Review No. 27*. United States of America: OECD Publishing. 129-146.

- DERNIS H, KHAN M. (2004). Triadic patent families methodology. En *OECD Science, Technology and Industry Working Papers, 2004/2*. United States of America: OECD Publishing. 1-33.
- EPO. (2012). *Member states of the European Patent Organisation*. Disponible en: <http://www.epo.org/about-us/organisation/member-states.html> [Consultado: 10 febrero de 2012].
- FAYYAD U, PIATESTSKY G, SMYTH P. (1996). From data mining to knowledge discovery in databases. *American Association for Artificial Intelligence*. 41(1): 37–54.
- FERNÁNDEZ M. (2001). Indicyt: indicadores de ciencia y tecnología en red. *Cuadernos de Indicios*. 1: 101–09.
- FERNÁNDEZ M *et al.* (1993). Relational databases: constructing a relational database for bibliometric analysis. *Research Evaluation*. 3(1): 55–62.
- FERNÁNDEZ M *et al.* (2001). Indicyt: science and technology indicators in Spain: development of an application for interactive search on the internet. *Research Evaluation*. 10(2): 83–88.
- FERNÁNDEZ MARCIAL V. (2006). Gestión del conocimiento vs. gestión de la información. *Investigación bibliotecológica*. 20(41): 44–62.
- FRAWLEY WJ, PIATETSKY-SHAPIRO G, MATHEUS CJ. (1992). Knowledge discovery in databases: an overview. *AI Magazine*. 13(3): 57–70.
- FRIETSCH R, SCHMOCH U. (2010). Transnational patents and international markets. *Scientometrics*. 82(1): 185–200.
- FRUCHTERMAN TM, REINGOLD EM. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*. 21(11): 1129–64.
- GARCÍA-ESCUADERO MÁRQUEZ P, LÓPEZ LÓPEZ P. (1997). Análisis bibliométrico y literatura de patentes. *Revista General de Información y Documentación*. 7(2): 181–199.
-

- GEPHI CONSORTIUM. (2011). The open graph viz platform, disponible en: <https://gephi.org/> [Consultado: 15 de marzo de 2011].
- GEPHI.(2011). Tutorial layouts, disponible en: <https://gephi.org/users/tutorial-layouts/> [Consultado: 15 de marzo de 2011].
- GLÄNZEL W, MEYER M. (2003). Patents cited in the scientific literature: an exploratory study of 'reverse' citation relations. *Scientometrics*. 58(2): 415–28.
- MYSQL. (2012). El sistema operativo GNU, disponible en: <http://www.gnu.org/licenses/licenses-list.html> [Consultado: 23 de mayo de 2012].
- GÓMEZ I *et al.* (2005). Typology for indicators. Regionalisation of science and technology in spain. *Reserch Evaluation*. 14(2): 137–48.
- GONZÁLEZ-ALBO MANGLANO B, ZULUETA MA. (2007). Estudio comparativo de bases de datos de patentes en internet. *Anales de Documentación*. (10): 145–62.
- GONZÁLEZ ALBO MANGLANO B, ZULUETA MA. (2007). Patentes domésticas de universidades españolas: análisis bibliométrico. *Revista Española de Documentación Científica*. 31(1): 61–90.
- GORBEA-PORTAL S. (1994). Principios teóricos y metodológicos de los estudios métricos de la información. *Investigación bibliotecológica*. 8(17): 23–32.
- GORBEA-PORTAL S. (2005). *Modelo Teórico para el Estudio Métrico de la Información Documental*. Gijón: TREA.
- GORBEA-PORTAL S. (2006). Perspectivas interdisciplinarias de los estudios métricos de la información. En *Memoria del XXIII Coloquio de Investigación Bibliotecológica y de la Información: problemas y métodos de investigación en Bibliotecología e Información. Una perspectiva interdisciplinaria*. México: Centro Universitario de Investigaciones Bibliotecológicas. p. 17–30.
- GORBEA-PORTAL S., Suárez-Balseiro, C.A. (2007). Análisis de la influencia y el impacto entre revistas periféricas no incluidas en el Science Citation Index. *Revista Interamericana de bibliotecológica*. 30(2): 47–70.

- GRILICHES Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature*. 28(4): 1661–1707.
- GROENINGER M. (2010). Circular layout, disponible en: <https://marketplace.gephi.org/plugin/circular-layout/> [Consultado: 20 de noviembre de 2010].
- GUANGZHOU HU A. (2009). The regionalization of knowledge flows in east asia: evidence from patent citations data. *World Development*. 37(9): 1465–77.
- GUASCH LM. (2007). Nuevas perspectivas en la evaluación de las patentes como parte del curriculum científico. *Revista Española de Documentación Científica*. 30(2): 218–40.
- GUENTHER K. (2000). Applying data mining principles to library data collection. *Computers in libraries*. 20(4): 60–63.
- GUZMÁN SÁNCHEZ MV. (1999). *Patentometría. Herramienta para el análisis de oportunidades tecnológicas* [tesis doctoral]. La Habana: Facultad de Economía. Universidad de La Habana.
- GUZMÁN SÁNCHEZ MV, SOTOLONGO AGUILAR G. (2002). Mapas tecnológicos para la estrategia empresarial. Situación tecnológica de la neisseria meningitidis. *Acimed*. 10(4): 25-30.
- HALL BH, JAFFE A, TRAJTENBERGL M. (2005). Market value and patent citations. *The Rand Journal of Economics*. 36(1): 16–38.
- HAMMERGREN T. (1996). *Data Warehousing. Bulding the corporate knowledge base*. United States of America: International Thompson Computer Press.
- HAN J, KAMBER M. (2001). *Data mining: concepts and techniques*. United States of America: Morgan Kaufmann Publishers.
- HARHOFF D *et al.* (1999). Citation frequency and the value of patented inventions. *The Review of Economics and Statistics*. 81(3): 511–15.

- HARHOFF D, SCHERER FM, VOPEL K. (2003). Citations, family size, opposition and the value of patent right. *Research Policy*. (32): 1343–1363.
- HERNÁNDEZ ORALLO J, RAMÍREZ QUINTANA M J, FERRI RAMÍREZ C. (2004). *Introducción a la minería de datos*. Madrid: Prentice Hall.
- HICKS D. (2000). 360 degree linkage analysis. *Research Evaluation*. 8(2): 133–143.
- HIDALGO-NUCHERA A, IGLESIAS-PRADAS S, HERNÁNDEZ-GARCÍA A. (2009). Utilización de las bases de datos de patentes como instrumento de vigilancia tecnológica. *El Profesional de la Información*. 18(5): 511–19.
- HINGLEY P. (2009). Patent families defined as priority forming filings and their descendents. En *EPO/OECD Patent families workshop. Working Papers, 2009/2*. United States of America: OECD Publishing; 2009. 1-48.
- HINGLEY P, NICOLAS M. (2004). Methods for forecasting numbers of patent applications at the european patent office. *World Patent Information*. (26): 191–204.
- HU Y. (2005). Efficient and high quality force-directed graph drawing. *The Mathematical Journal*. 10: 37–71.
- HUDOMALI E, VIDMAR G. (2003). Olap and bibliographic databases. *Scientometrics*. 58(3): 609–22.
- HULLMANN A, MEYER M. (2003). Publications and patents in nanotechnology. *Scientometrics*. 58(3): 507–527.
- HURD J. (2000). The transformation of scientific communication: a model for 2020. *Journal of the American Society for Information Science*. 51(14): 1279–83.
- IGAMI M. (2008). Exploration of the evolution of nanotechnology via mapping of patent applications. *Scientometrics*. 77(2): 289–308.
- INMON W, STRAUSS D, NEUSHLOSS G. (2008). *DW 2.0. The architecture for the next generation of data warehousing*. United States of America: Morgan Kaufmann.

- JOHNSON D. (2004). A data mining primer and implications for school library media specialists. *Knowledge Quest*. 32(5): 32–35.
- JUNQUERA B, MITRE M. (2007). Value of bibliometric analysis for research policy: a case study of spanish research into innovation and technology management. *Scientometrics*. 71(3): 443–54.
- JACOMY M *et al.* (2011). ForceAtlas2, a graph layout algorithm for handy network visualization, disponible en: http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf/ [Consultado: 15 de marzo de 2011].
- KATZ J, HICKS D. (1997). Desktop scientometrics. *Scientometrics*. 38(1): 141–53.
- KLITKOU A, GULBRANDSEN M. (2010). The relationship between academic patenting and scientific publishing in Norway. *Scientometrics*. 82(1): 93–108.
- KÜRTÖSSY J. (2004). Innovation indicators derived from patent data. *Periodica Polytechnica Ser. Soc. Man. Sci.* 12(1): 91–101.
- LAMPE R. (2012). Strategic citation. *World Development*. 2012; 37(9): 1465– 77.
- LASCURAIN LM *et al.* (2010). Capacidad innovadora de la Comunidad de Madrid a partir de las patentes concedidas entre 1996-2007. *Revista Española de Documentación Científica*. 33(3): 458–79.
- LASCURAIN ML, MADERA JARAMILLO MJ, SANZ CASADO E. (2009). La capacidad innovadora de la universidad española frente a su actividad publicadora. En *Evaluación de la calidad de la educación superior y la investigación (VI Foro)*. España: Asociación Española de Psicología Conductual. p. 118
- LASCURAIN ML, MADERA JARAMILLO MJ, SANZ CASADO E. (2010). Origen del conocimiento tecnológico de las patentes españolas. En *Evaluación de la calidad de la educación superior y la investigación (VII Foro)*. España: Asociación Española de Psicología Conductual. p. 118
- LASCURAIN ML, MADERA JARAMILLO MJ, SANZ CASADO E. (2011). Interés de las invenciones españolas a partir del estudio de las familias de patentes. En

- Evaluación de la calidad de la educación superior y la investigación (VIII Foro)*. España: Asociación Española de Psicología Conductual. p. 299
- LEE YG. (2010). Sectoral strategic differences of technological development between electronics and chemistry: a historical view from analyses of koreaninvented us patents during the period of 1989-1992. *The Journal of Information and Knowledge Management Systems*. 81(1): 83–92.
- LEYDESDORFF L. (2008). Patent classifications as indicators of intellectual organization. *Journal of the American Society for Information Science and Technology*. 59(10): 1582–97.
- LEYDESDORFF L, MEYER M. (2007). The scientometrics of a triple helix of university-industry-government relations (introduction to the topical issue). *Scientometrics*. 70(2): 207–22.
- LEYDESDORFF L, WAGNER C. (2009). Macro-level indicators of the relations between research funding and research output. *Journal of Informetrics*. (3): 353–62.
- LI Z. (2009). Developing a systematic patent search training program. *The Journal of Academic Librarianship*. 35(3): 260–66.
- MAGERMAN T, LOOY BV, SONG X. (2010). Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*. 82(2): 289–306.
- MARTIN S *et al.* (2011). OpenOrd: an open-source toolbox for large graph layout. En *SPIE Conference on Visualization and Data Analysis (VDA)*. Disponible en: http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDoQFjAB&url=http%3A%2F%2Fwww.researchgate.net%2Fpublication%2F253087985_OpenOrd_an_open-source_toolbox_for_large_graph_layout%2Ffile%2F3deec5205279e8c66a.pdf&ei=rfSQU7jyE4mSqAaZ8oLwAw&usg=AFQjCNGTiVrjyHdlT3u3TkETNGirxbGgQ&sig2=QIPa3oN4gOt8-Alqg5-TjQ&bvm=bv.68445247,d.b2k [Consultado: 10 de marzo de 2012].

- MARTÍNEZ C. (2010). Insight into different types of patent families. En *OECD Science, Technology and Industry Working Papers*, 2010/2. United States of America: OECD Publishing. 1-49.
- MARTÍNEZ C. (2011). Patent families: When do different definitions really matter? *Scientometrics*. 86(1): 39–63.
- MARTÍNEZ-MÉNDEZ FJ, PASTOR-SÁNCHEZ JA, LÓPEZ-CARREÑO R. (2010). Las patentes como indicador de la actividad científica en las universidades españolas. *El Profesional de la Información*. 19(2): 168–74.
- MCMILLAN GS, NARIN F, DEEDS DL. (2000). An analysis of the critical role of public science in innovation: the case of biotechnology. *Research Policy*. (29): 1–8.
- MEYER M. (2000). Does science push technology? patents citing scientific literature. *Research Policy*. (29): 409–34.
- MEYER M. (2000). What is special about patent citations? differences between scientific and patent citations. *Scientometrics*. 49(1): 93–123.
- MEYER M. (2001). Patent citation analysis. a closer look at the basic input data from patent search reports. *Scientometrics*. 51(1): 185–201.
- MEYER M. (2001). Patent citation analysis in a novel field of technology: an exploration of nano-science and nano-technology. *Scientometrics*. 51(1): 163–183.
- MEYER M. (2002). Tracing knowledge flow in innovation systems. *Scientometrics*. 54(2): 193–212.
- MEYER M. (2006). Measuring science-technology interaction in the knowledge-driven economy: The case of a small economy. *Scientometrics*. 66(2): 425–39.
- MEYER M, BHATTACHARYA S. (2004). Commonalities and differences between scholarly and technical collaboration. An exploration of co-inventors and coauthorship analysis. *Scientometrics*. 61(3): 443–56.

Referencias

- MEYER M *et al.* (2005). Inventive output of academic research: a comparison of two science systems. *Scientometrics*. 63(1): 145–161.
- MIELNICZUK DE MOURA AM. (2009). *A Interação entre artigos e patentes: um estudo cientométrico da comunicação científica e tecnológica em biotecnologia* [tesis doctoral]. Brasil: Universidade Federal do Rio Grande do Sul, Porto Alegre.
- MYSQL. (2012). Información general, disponible en: <http://dev.mysql.com/doc/refman/5.0/es/introduction.htmk> [Consultado: 23 de mayo de 2012].
- NARIN F. (1994). Patents bibliometrics. *Scientometrics*. 30(1): 147–55.
- NARIN F *et al.* (1997). The increasing linkage between u.s. technology and public science. *Research Policy*. (26): 317–30.
- NARIN F, OLIVASTRO D. (1993). Patent citation cycles. *Library Trends*. 41(4): 700–09.
- NOOY W, MRVAR A, BATAGELJ V. (2011). *Exploratory social network analysis with Pajek*. United States of America: Cambridge University Press.
- OCDE. (2002). *Manual de Frascati. Propuesta de norma práctica para encuestas de investigación y desarrollo experimental*. 3a ed. España: FECYT, OCDE.
- OCDE. (2005). *Manual de Oslo. Guía para la recogida e interpretación de datos sobre innovación*. 3a ed. España: OCDE, European Communities, TRAGSA.
- OCDE. (2009). *Manual de estadísticas de patentes de la OCDE*. Madrid: Oficina Española de Patentes y Marcas (OEPM).
- OMPI. (2005) Norma ST.3. Códigos normalizados de dos letras, recomendaciones para la representación de estados, otras entidades y organizaciones intergubernamentales. En *Manual de información y documentación en materia de propiedad industrial*. [s.l.]: OMPI.
- OPPENHEIM C. (1997). Patent citation analysis. *Scientometrics*. 39(1): 141.

- PAKES A, SCHANKERMAN M. (1984). *R & D, Patents, and Productivity*. En *The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources*. United States of America: University of Chicago Press. p. 73–88
- PAO-LONG C, CHAO-CHAN W, HOANG-JYH L. (2010). Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics*. 82(1): 5–19.
- PARK H W, KANG J. (2009). Patterns of scientific and technological knowledge flows based on scientific papers and patents. *Scientometrics*. 81(3): 811–20.
- PATENTE LENS. (2012). What is a continuation application? Disponible en: <http://www.patentlens.net/daisy/patentlens/2645.html> [Consultado: 10 de marzo de 2012].
- PLAZA LM, ALBERT A. (2004). Análisis de la producción científica española citada en patentes biotecnológicas en EE. UU. *Revista Española de Documentación Científica*. 27(2): 212–20.
- PLAZA LM, ALBERT A. (2008). Scientific literature cited in uspto patent documents as indicators for the evaluation and analysis of spanish scientific research in biomedical disciplines. *Scientometrics*. 76(3): 429–38.
- POTTELSBERGHE DE LA POTTERIE BV, ZEEBROECK NV. (2008). A brief history of space and time: the scope-year index as a patent value indicator based on families and renewals. *Scientometrics*. 75(2): 319–338.
- RAMOS-VIELBA I, FERNÁNDEZ-ESQUINAS M, ESPINOSA-DE-LOS MONTERO E. (2010). Measuring university-industry collaboration in a regional innovation system. *Scientometrics*. 84(3): 649–67.
- ROSELL C, AGRAWAL A. (2009). Have university knowledge flows narrowed? Evidence from patent data. *Research Policy*. (38): 1–13.

- RUSSELL J, ROUSSEAU R. (2002). *Bibliometrics and Institutional Evaluations*. En Encyclopedia of Life Support Systems (EOLSS), part 19.3 Science and Technology Policy. United Kingdom: Developed under the auspices of the UNESCO, Eolss Publishers, Oxford.
- SANCHO R. (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica. *Revista Española de Documentación Científica*. 13(3-4): 842–65.
- SCHWARTZ RL, CHRISTIANSEN T. (1997). *Learning Perl*. 2a ed. United States of America: O'reilly and Associates.
- SIMMONS, ES. (2009). "black sheep" in the patent family. *World Patent Information*. (31): 11–18.
- SOTOLONGO-AGUILAR G *et al.* (2001). Mining informetric data with self-organizing maps. En *Proceedings of the 8th International Society for Scientometrics and Informetrics*. Australia: Bibliometric & Informetric Research Group (BIRG), UNSW. p. 665–73
- SOTOLONGO-AGUILAR G, GUZMÁN-SÁNCHEZ M. (2001). Aplicaciones de las redes neuronales. El caso de la bibliometría. *Ciencias de la Información*. 32(1): 27–34.
- SOTOLONGO-AGUILAR G, GUZMÁN-SÁNCHEZ M, CARRILLO H. (2001). Vibliosom: visualización de información bibliométrica mediante el mapeo autoorganizado. *Revista Española de Documentación Científica*. 25(4): 477–84.
- SOTOLONGO-AGUILAR G, SUÁREZ-BALSEIRO C, GUZMÁN-SÁNCHEZ M. (2000). Modular bibliometric information system with proprietary software (mobisprosoft): a versatile approach to bibliometric research tools. *Library and Information Science Electronic Journal*. 10(2): 35-48.
- SRINIVANSAN S. (1997). *Advances Perl Programming*. United States of America: O'reilly and Associates.

- STARESINIC M, BOH B. (2009). The issue of relevance in full-text patent document searches. *Patent Informatics*. 33(1): 175–72.
- STERN D. (2003). New knowledge management systems: the implications for data discovery, collection development, and the changing role of the librarian. *Journal of the American Society for information science and technology*. 54(12): 1138–40.
- STERNITZKE C. (2009). Defining triadic patent families as a measure of technological strength. *Scientometrics*. 81(1): 91–109.
- STERNITZKE C. (2009). Reducing uncertainty in the patent application procedure insights from invalidating prior art in European patent applications. *World Patent Information*. 2009; (31): 48–53.
- SZU-CHIA SL. (2010). Scientific linkage of science research and technology development: a case of genetic engineering research. *Scientometrics*. 82(1): 109–120.
- TAPPEINER G, HAUSER C, WALDE J. (2008). Regional knowledge spillovers: Fact or artifact? *Research Policy*. (37): 861–74.
- THOMSON C. (2006). *Derwent. World patents index. Dialog online user guide*. 3a ed. United Kingdom: Thomson Scientific.
- THOMSON C. (2004). *Derwent Innovation Index 4.0*. Disponible en: http://biblioteca.unizar.es/guia_bd/dii4_sem_0104_es.pdf [Consultado: 15 enero de 2009].
- VAN ECK NJ, WALTMAN L. (2009). How to normalize cooccurrence data? an analysis of some well-know similarity measures. *Journal of the American Society for Information Science and Technology*. 60(8): 1635– 51.
- VAN ECK NJ, WALTMAN L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 84(2): 523– 38.

- VAN ECK NJ *et al.* A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology.*, 2010; 61(12): 2405–16.
- VAN ECK NJ, WALTMAN L, NAYONS E. C. (2010). Automatic term identification for bibliometric mapping. *Scientometrics.* 2010; 82(3): 582–596.
- VAN ECK NJ. (2011). *Methodological advances in bibliometric mapping of science* [tesis doctoral]. Netherlands: Erasmus Research Institute, Erasmus University Rotterdam.
- VAN ECK NJ, WALTMAN L. (2011a). Text mining and visualization using VOSviewer. *ISSI Newsletter.* 7(3): 50–54.
- VAN ECK NJ, WALTMAN L. (2011b). *Vosviewer manual. Informe Técnico Manual for VOSviewer version 1.4.0.* Netherlands: Universiteit Leiden.
- VAN ZEEBROECK N, VAN POTTELSBERGHE DE LA POTTERIE B. (2007). Filing strategies and patent value. En *Segunda conferencia anual de la European Policy for Intellectual Property (EPIP)*. Sweden: Lund. 1-34.
- VILELA CHAVES C, MORO S. (2007). Investigating the interaction and mutual dependence between science and technology. *Research Policy.* (36): 1204–1220.
- WALL L, CHRISTIANSEN T, SCHWARTZ RL. (1996). *Programming Perl.* 2a ed. United States of America: O'reilly and Associates.
- WALTMAN L, VAN ECK NJ, NAYONS EC. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics.* 4(4): 629–35.
- WANNER L *et al.* (2008). Towards content-oriented patent document processing. *World Patent Information.* (30): 21– 33.
- WEBB C *et al.* (2005). Analysing European and international patent citations: a set of epo patent database building blocks. En *OECD Science, Technology and Industry Working Papers, 2005/9.* United States of America: OECD Publishing.

- WIPO. (2008). World patent report. A statistical review. Informe Técnico WIPO Publication No 931(E). [s.l.]: World Intellectual Property Organization.
- WIPO. (2011a). *Convenio de París para la protección de la propiedad Industrial (enmendado el 28 de septiembre de 1979)*, disponible en: http://www.wipo.int/treaties/es/text.jsp?file_id=287557 [Consultado: 20 marzo de 2011].
- WIPO. (2011b). *Introducción al Tratado de Cooperación en materia de Patentes*. (Material recibido en el Curso de Enseñanza a Distancia sobre el PCT).
- WITTEN I *et al.* (2004). Text mining in a digital library. *International Journal Digital libraries*. 4(1): 56–59.
- YANG YY *et al.* (2008). Text mining and visualization tools -impressions of emerging capabilities. *World Patent Information*. (30): 280–93.

ANEXO

Anexo

Estados miembro de la Oficina Europea de Patentes

Código	Nombre del país	Miembro desde
AL	Albania	1 May 2010
AT	Austria	1 May 1979
BE	Belgium	7 October 1977
BG	Bulgaria	1 July 2002
CH	Switzerland	7 October 1977
CY	Cyprus	1 April 1998
CZ	Czech Republic	1 July 2002
DE	Germany	7 October 1977
DK	Denmark	1 January 1990
EE	Estonia	1 July 2002
ES	Spain	1 October 1986
FI	Finland	1 March 1996
FR	France	7 October 1977
GB	United Kingdom	7 October 1977
GR	Greece	1 October 1986
HR	Croatia	1 January 2008
HU	Hungary	1 January 2003
IE	Ireland	1 August 1992
IS	Iceland	1 November 2004
IT	Italy	1 December 1978
LI	Liechtenstein	1 April 1980
LT	Lithuania	1 December 2004
LU	Luxembourg	7 October 1977
LV	Latvia	1 July 2005
MC	Monaco	1 December 1991
MK	Former Yugoslav Republic of Macedonia	1 January 2009
MT	Malta	1 March 2007
NL	Netherlands	7 October 1977
NO	Norway	1 January 2008
PL	Poland	1 March 2004
PT	Portugal	1 January 1992
RO	Romania	1 March 2003
RS	Serbia	1 October 2010
SE	Sweden	1 May 1978
SI	Slovenia	1 December 2002
SK	Slovakia	1 July 2002
SM	San Marino	1 July 2009
TR	Turkey	1 November 2000

Estados que reconocen las patentes europeas bajo petición

Código	Nombre del país
BA	Bosnia and Herzegovina
ME	Montenegro